Technology Service Corporation

2811 WILSHIRE BOULEVARD • SANTA MONICA, CALIFORNIA 90403 • PH. (213) 829-7411

TSC-PD-B611-3

Final Report for DEVELOPMENT OF DATA BASE FOR HUMAN EXPOSURE TO AIR POLLUTION IN THE SOUTH COAST AIR BASIN

Contract A7-163-30

October 1979

by

Yuji Horie
Dennis Bicker

Submitted to:

California Air Resources Board
Research Division
P.O. Box 2815
Sacramento, California 95812

Project Officer: Dane Westerdahl

TABLE OF CONTENTS

1. INTRODUCTION AND SUMMARY

The purpose of this project is to develop a common data base of human exposure to air pollution in the South Coast Air Basin (SCAB), so that the same data base can be used for various epidemiological studies of air pollution health effects. In the past studies, the level of each individual's exposure to air pollution was assumed to be the same as the exposure level of the community from which a controlled (i.e., study) population was selected. However, as an air pollution level varies with time and space, so does an exposure level of a person, depending upon his spatial position, daily activity pattern, and other factors.

Some health effects are considered to be caused more by a short-term peak (acute) exposure rather than by a long-term average (chronic) exposure. The opposite may be true for some other health effects. Therefore, to perform a comprehensive epidemiological study requires both short-term and long-term exposure estimates. People are exposed simultaneously to several pollutants. Simultaneous exposures to multiple pollutants may cause either synergistic (more than additive) or antagonistic (less than additive) effects on human health. Therefore, the developed data base contains human exposure data for each of the six major (four gaseous and two particulate) criteria pollutants. To help investigate those health effects which have a long latency period, the data base covers an extended period of eleven years, from January 1966 through December 1976.

1.1 Outline of Data Base

A large human exposure data base has been developed to provide a convenient human exposure data source for performing an epidemiological study of air pollution health effects. The data base includes monthly values of several key exposure parameters, computed for each of the four gaseous (OX, NO2, SO2 and CO) and two particulate (TSP and SO4) pollutants at some 400 postal zip-code areas as well as at some 40 air monitoring stations scattered over the SCAB (see Figure 1 and Attachment F). Hourly concentration data were used to compute exposure values for the gaseous pollutants, while 24-hour concentration data were used for the particulate pollutants.

Given the residence location of a person by zip-code, one can readily extract from the data base necessary information about outdoor exposure characteristics of that person for any month or period during the 11-year period. Given UTM coordinates of a person or a group of people, one can easily compute from the data base monthly exposure values by applying the attached spatial interpolation program to the human exposure data at air monitoring sites. Given a history of a person's residences and/or work places, one can compute from the data base both cumulative exposure values and a history of exposures by month.

A certain adverse health effect may be caused either by a short-term exposure to peak concentrations or by a long-term exposure to low concentrations. To cope with such uncertainty in the air pollution-health effects relationship, the exposure data are computed for both total exposures and excess exposures. Here, an excess exposure means that a person is exposed to pollutant concentrations above a given concentration threshold. For each pollutant, four different concentration thresholds were chosen by considering the known (but not quantified) health effects from exposures to that pollutant.

The total exposure is given each month by a mean concentration and an apparent dose (i.e., time integral of concentration). The excess exposure is given by an exceedance frequency (i.e., number of hours exceeding a given threshold) and an excess dose (i.e., time integral of concentrations exceeding a threshold level). Both the total and excess exposure data are computed each month for each pollutant at both the monitor site and the centroid of each postal zip-code area. Exposure data at each zip-code area are labeled by a confidence indicator, which was determined by computing the distance between the centroid of that zip-code area and the nearest monitoring station with valid air quality data. If the distance is less than a

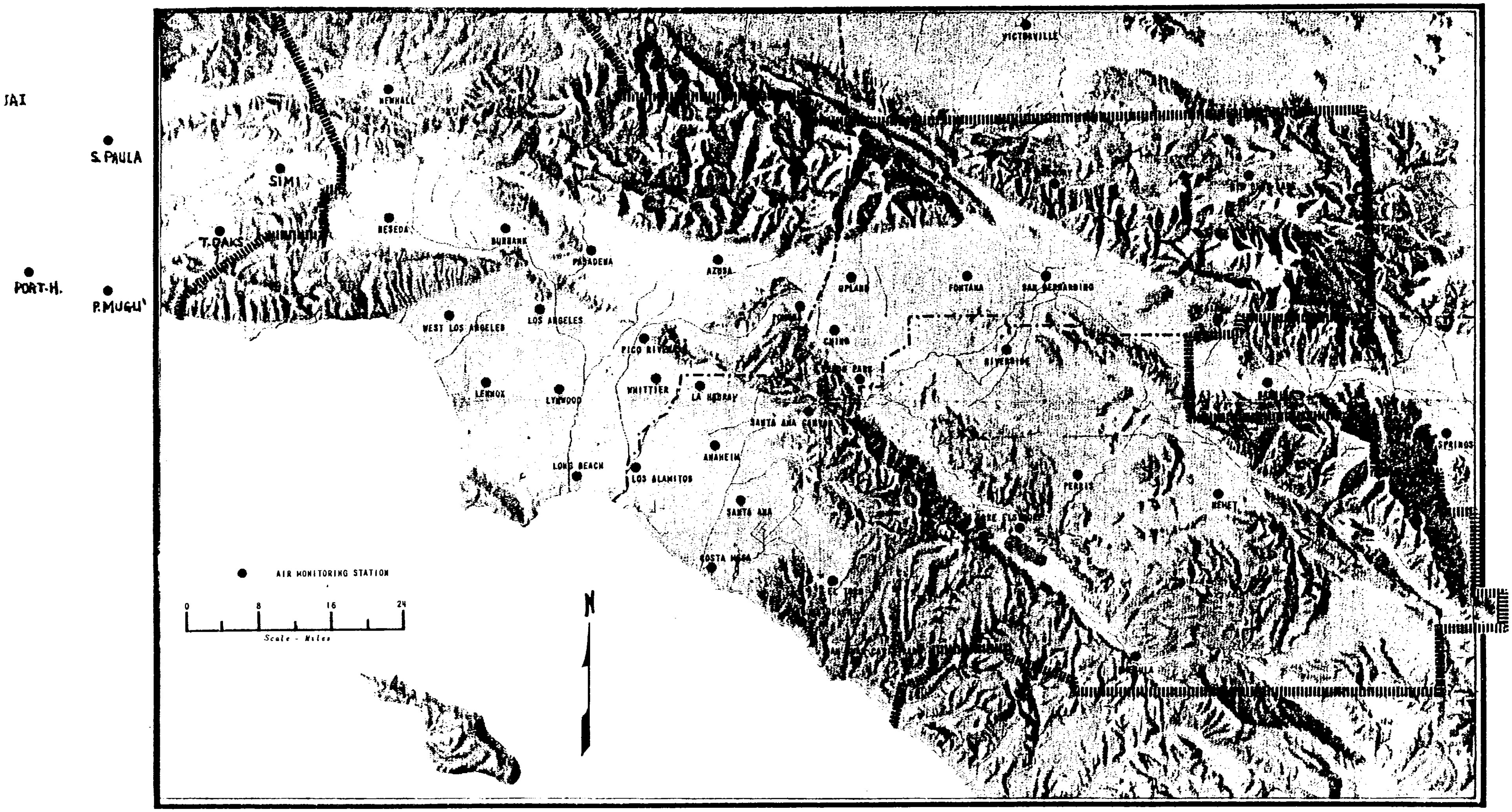


Figure 1. Location of Air Monitoring Stations Used for Developing Human Exposure Data Bases in SCAB Region

typical representative area of a monitor for a given pollutant, the confidence indicator for that zip-code area and pollutant is denoted as A, the highest confidence level. If the distance is greater than the typical representative area but less than twice the typical representative area, the confidence indicator, B, is used to indicate the moderate confidence in the data. If the distance is greater than twice the typical representative area, C is used to indicate that the exposure data have the least confidence.

1.2 Recommended Use of the Data Base

The developed data base provides spatial and temporal characteristics of apparent exposure, i.e., that estimated from ambient air quality monitoring data, instead of true exposure to air pollution. Therefore, if one suspects that exposure to a certain pollutant occurs mostly indoors rather than outdoors, the data base should not be used for such a case. However, if human exposure occurs partially indoors and partially outdoors, the data base can be used to approximate the outdoor exposure. In most epidemiological studies, there is no praticable means of measuring the indoor exposure. It is recommended that a substitutable quantity be found from information obtainable from survey questionnaires about daily activity patterns and housing characteristics.

Given a person's residence history, his cumulative exposure can be computed easily by retrieving from the data base a set of exposure data for his residence in Period P1 and Zip-Code Area Z1, another set of exposure data in Period P2 and Zip-Code Area Z2, and so on. All we have to do is just sum those exposure values over the entire study period. If one wants to know seasonal exposure, he can add the monthly exposure values for those months that belong to the same season.

If one would like to refine the exposure estimate for a person whose residence location and work place are both known, the following procedure is recommended: First, all day exposures at the person's residence and work place are computed from the data base. Second, the exposure value at his residence is adjusted for weekend/weekday air quality differences by multiplying it by 0.93 for photochemical oxidants, 1.0 for sulfate, and 0.75 for the other four pollutants [Horie et al. 1979]. Then, the person's exposure is estimated by taking a weighted average of the two exposure values, i.e.,

$$[5(\text{exposure at work place}) + 2(\text{exposure at residence})] \div 7 \quad .$$

The above procedure of refining an exposure estimate will work rather well for photochemical oxidants, whose nighttime concentrations are virtually zero. For other pollutants whose nightime levels can be high, the above procedure will not work as well as for photochemical oxidants. The reason is that the recommended procedure does not take into account a difference between nighttime exposure at the residence location and that at the work place. To correct the defect of the recommended procedure may require a further refinement of indoor/outdoor exposures, which is beyond the scope of the present study.

In some cases, one would like to have exposure parameter values for a threshold that is different from those used in the data base. In such cases, one can use either the percentile concentration data tabulated for the gaseous pollutants or the original concentration data tabulated for the particulate pollutants. By applying the following human-exposure program to the percentile or original concentration data, one can compute exposure parameter values for a desired concentration threshold. In some other cases, one may want to have exposure data at regularly spaced grid points instead of at the centroid of each zip-code area. In such cases, one can use the exposure data computed at each air monitoring station and the spatial interpolation computer program attached to this data base. All he has to do is to apply the spatial interpolation program to the station exposure data and compute the desired exposure data at each grid point.

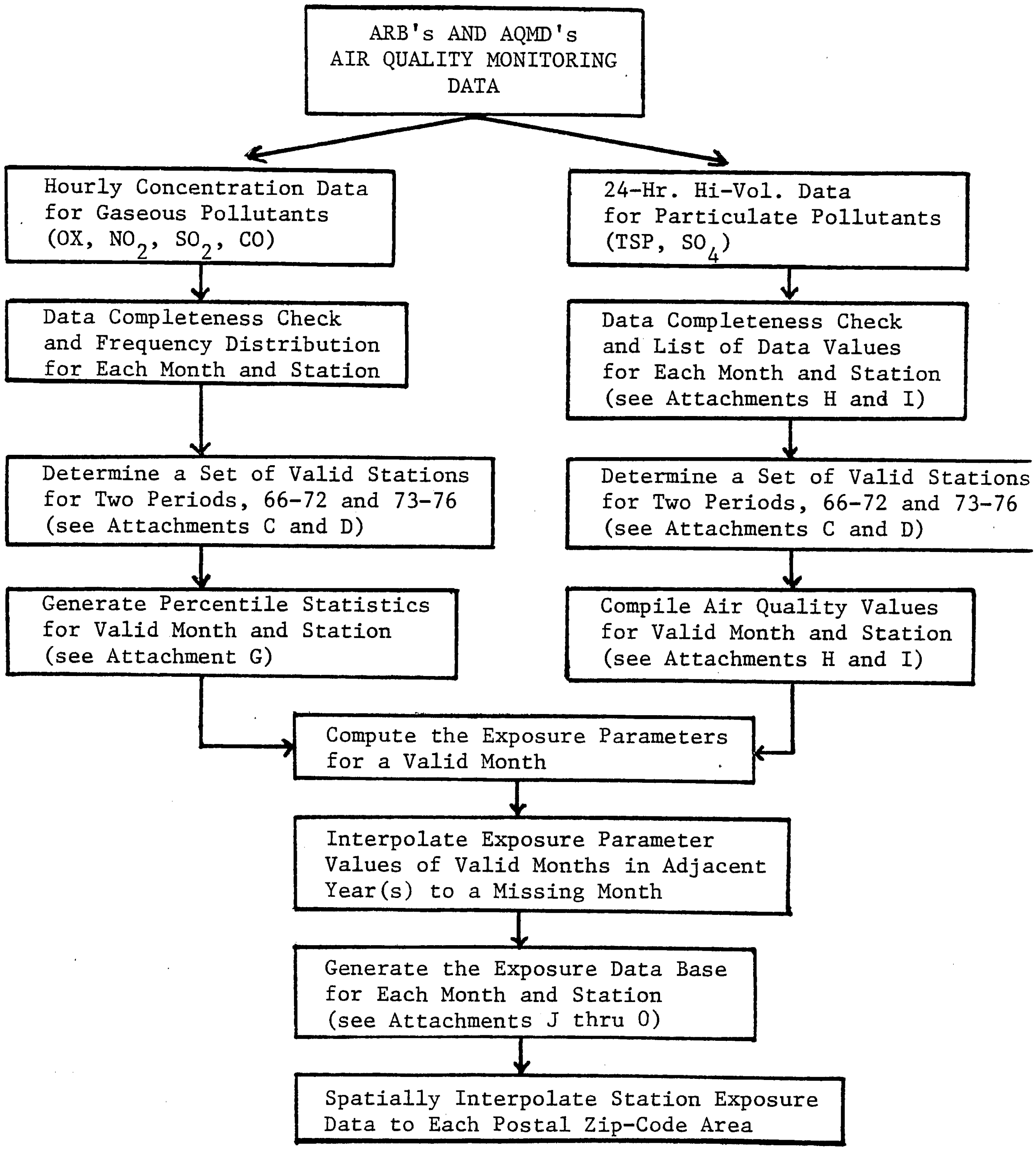


Figure 2. Flow Diagram of Data Processing Steps from Original Air Quality Data to Human Exposure Data (see Attachment A)

Table 1. Number of Valid Stations* Used for Developing Human Exposure Data during the 1st (1966-72) and 2nd (1973-76) Periods.

Pollutant	1966-1972	1973-1976
Photochemical Oxidants (OX)	11	30
Nitrogen Dioxide (NO2)	8	20
Carbon Monoxide (CO)	11	26
Sulfur Dioxide (SO2)	2	17
Total Suspended Particulate Matter (TSP)	0	27
Sulfate (SO4)	0	7

*For valid stations, see Attachments C and D.

The criteria used to select a valid station for gaseous pollutants are as follows:

1. The station must report air quality data for at least 50% of the total number of hours per month to be a valid station for that month.
2. The station must have at least 7 valid months per year to be a valid station for that year.
3. The station must be valid in 5 out of 7 years during the 1st period and in 3 out of 4 years during the 2nd period to be a valid station.

The criteria used to select a valid station for a particulate pollutant are similar to those for a gaseous pollutant. However, the first criterion is changed to the following:

1. A station must report air quality data for at least two days per month to be valid for that month.

Valid stations selected according to the criteria described above are listed in Attachment C for the first period, 1966-1972, and in Attachment D for the second period, 1973-1976.

2.2 Development of Air Quality Data Base

When a human exposure data base is to be developed from ambient air quality monitoring data, a number of problems arise: First, the number of monitoring stations that report an adequate number of hourly or daily air quality values varies from year to year. If a changing set of monitoring stations is used to estimate human exposure levels at a non-station location, the reliability and accuracy of the estimated exposure values may vary from year to year. Therefore, the same set of valid monitoring stations was used to compute human exposure levels at each postal zip-code area for each pollutant and for each of the two study periods, 1966-1972 and 1973-1976.

Second, air quality data for a gaseous pollutant are too voluminous for simple computation of monthly exposure values. Therefore, the gaseous pollutant data were compressed by computing a cumulative distribution of hourly concentrations for each month at each station. The cumulative distribution was approximated by computing concentration values at selected percentiles: every .25% above 99% (four values), every 1% above 90% (nine), every 2.5% above 70% (eight), every 5% above 50% (four), every 10% equal to or above 0% (six). A concentration value at the 100th percentile was set equal to the maximum measured concentration, while that at the 0th percentile was set equal to the minimum measured concentration. In addition to the above percentile values, an arithmetic mean concentration was also computed each month for each station. By converting hourly concentration data into the percentile form described above, we compressed the air monitoring data by about 22 to 1 (see Attachment G).

To compute the monthly percentile statistics, the original hourly concentration readings were sorted in a descending order. The highest concentration was ranked as r = 1, the second highest as r = 2, and so forth. For gaseous pollutant data, the percentile, P, of each concentration value was computed by the following approximate formula:

$$P = 1 - \frac{r}{N+1} \qquad (1)$$

where r is the rank of that concentration value and N is the sample size of the gaseous pollutant concentration data for a given month and station.

Third, air quality data for a particulate pollutant are too few to construct a believable cumulative distribution. While the cumulative distribution is to be specified by concentration values at the 31 (4+9+ 8+4+6) selected percentiles, the total number of particulate concentrations reported each month by each monitor seldom exceeds 6 (30 days/every 5th day). Therefore, the particulate concentration data are listed as they are, only by removing the dates on which the samples were taken (see Attachments H and I). In addition to the raw concentration data, the number of samples per month, the maximum and minimum values, and the geometric mean and standard deviation were computed for each month and station.

2.3 Exposure Data Base at Monitoring Sites

From the air quality data compiled each month for each station, monthly statistics of various exposure parameters were computed. The exposure parameters used in the current data base are as follows:

Symbol	Exposure Parameter	Unit
MEAN	Mean Exposure Level (a mean pollutant concentration over a month)	[Conc]
DOSE	Total Exposre (a cumulative exposure to air pollution over a month)	[Conc-Hr]
F	Exceedance Frequency (the number of hours exposed to concentrations above a given threshold over a month)	[Hr]
E	Excess Exposure (a cumulative exposure to concentrations above a given threshold over a month)	[Conc-Hr]

The mean exposure level (MEAN) at a monitoring site is approximated by a mean concentration level measured at that site. To make it consistent with the air quality standards, the mean exposure level is given by an arithmetic mean of hourly concentrations for the gaseous pollutants (OX, NO2, SO2 and CO) and by a geometric mean of 24-hour concentrations for the particulate pollutants (TSP and SO4).

The total exposure (DOSE) is given by a time integral of concentrations over a month. The total exposure can also be computed by multiplying the mean exposure level by the number of hours in a given month, i.e.,

$$\mathrm{DOSE} = \mathrm{MEAN} \times T \tag{2}$$

where T is the number of hours in that month.

A mean exposure level and a total exposure have been used for various epidemiological studies [Shy et al. 1970; Lave and Seskin 1977] to indicate a different level of human exposure for each sub-group of the study population. The two exposure parameters, DOSE and MEAN, may be a good exposure indicator if adverse air pollution health effects occur progressively from very low exposures to high exposures. However, an air quality standard has been set based on the premise that there exists a threshold concentration below which the public may never experience any adverse health effect from exposures to air pollution.

To facilitate future investigations of the above premise, the current human exposure data base includes two additional exposure parameters: Exceedance Frequency, F, and Excess Exposure, E. These parameter values are computed for each of the four threshold levels, which are listed in Attachment B for each of the six pollutants under study. As illustrated in Figure 3, the exceedance frequency and the excess exposure are given, respectively, by

$$F = T \times P_o \tag{3}$$

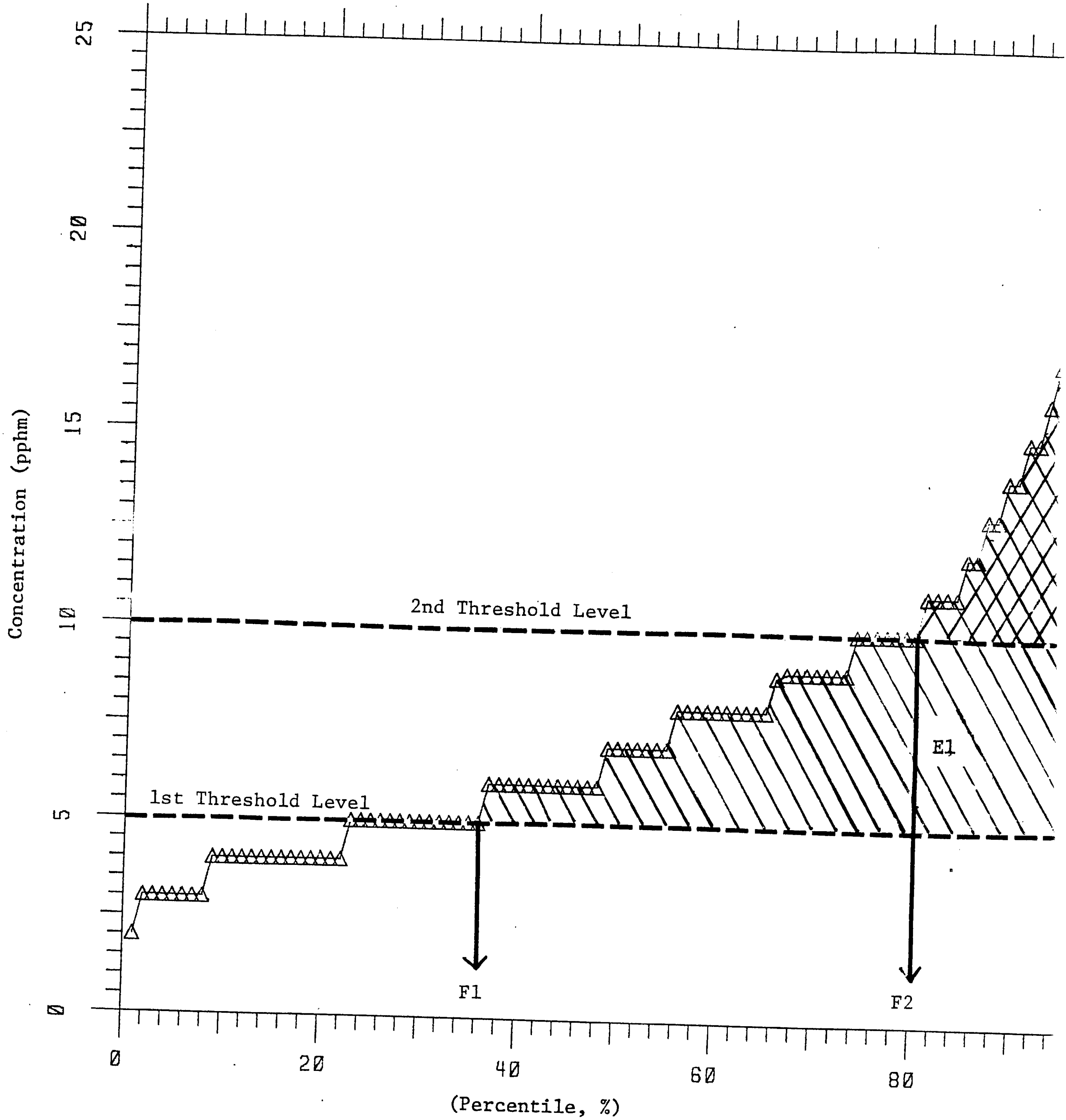


Figure 3. Cumulative Distribution of Hourly NO_2 Concentrations at Downtown Los Angeles for January 1978. (Excess exposures and exceedance frequencies for the 1st and 2nd threshold levels are given, respectively, by E1=hatched area x 744 hr, E2=double hatched area x 744 hr, F1=.36 x 744 hr, and F2=.80 x 744 hr.)

$$E = T \times \int_{P_o}^{100} [C(P) - C_s] \, dP \qquad (4)$$

where P_o is the percentile at which a percentile concentration, $C(P)$ becomes equal to a given threshold level, C_s.

Threshold Levels

For the last two exposure parameters, E and F, the parameter values are computed for each of four selected threshold levels as listed in Attachment B. For each pollutant, the threshold levels were chosen by considering actually occurring ambient levels and established concentration levels such as those used in the air quality standard and the health alert level. For example, an ambient level of daily maximum 1-hour oxidant concentrations varies from about 5 pphm to 50 pphm in the South Coast Air Basin (SCAB). The established oxidant concentration levels are 8 pphm (old national standard), 10 pphm (California standard), 12 pphm (new national standard), 20 pphm (1st stage alert level), and 35 pphm (2nd stage alert level). Therefore, we select the following threshold levels: 10 pphm for the 1st level; 15 pphm for the 2nd level; 20 pphm for the 3rd level; and 25 pphm for the 4th level.

Similarly, the thresholds for NO2 were selected as 5 pphm (national primary annual standard), 15 pphm, 20 pphm, and 25 pphm (California 1-hr standard). These threshold levels are well within the normal range of daily maximum 1-hr NO2 concentrations occurring in the SCAB region. The thresholds chosen for SO2 are 2, 4 (California 24-hr standard), 8, and 14 pphm (national primary 24-hr standard); those for CO are 5, 10 (California 12-hr standard), 15, and 20 pphm.

For TSP, there are California and national standards for both 24-hr concentrations and annual geometric mean concentrations. Therefore, the thresholds for TSP were mostly selected from the levels designated by those standards: 60 (California annual standard), 100 (California 24-hr standard), 150 (national secondary annual standard) and $200 \mu g/m^3$. On the other hand, there is only one air quality standard for SO4. This California 24-hr standard ($25 \mu g/m^3$) is violated rather infrequently in the SCAB region. Therefore, the threshold levels were chosen as 10, 15, 20, and 25 $\mu g/m^3$.

Filling in Missing Data

Although rather rigid criteria were employed to select valid stations for this data base development, some valid stations had several invalid months, i.e., months with less-than-complete data during each of the two study periods: January 1966 through December 1972, and January 1973 through December 1976. In estimating values of each exposure parameter for those invalid months, we tried three different methods of filling in the missing values.

i) Determine a typical seasonal pattern from J years of good data and then estimate a missing monthly value, say, the ith monthly average in the jth year, $X_{i,j}$ by

$$X_{i,j} = M_i \, Y_j \tag{5}$$

where M_i is the adjustment factor for the ith month and Y_j is the annual average of the jth year.

ii) Fill in the missing values by averaging the corresponding monthly values in two adjacent years, i.e., take

$$X_{i,j} = (X_{i,j-1} + X_{i,j+1}) \,/2 \tag{6}$$

iii) Fill in the missing value by substituting with the corresponding monthly value in one adjacent year, i.e., take

$$\left.\begin{array}{l} X_{i,j} = X_{i,j-1} \\ \text{or} \\ X_{i,j} = X_{i,j+1} \end{array}\right\} \qquad (7)$$

Initially we thought that the first method would yield a better estimate for a missing monthly value than the second and third methods. However, our trial results of each estimation method have indicated that the second and third methods often yield a better estimate for the missing monthly values than does the first method. In this exercise, we treated some valid months as missing months and compared the estimated values with the actual values to see the performances of the three estimation methods.

The above finding led to a more rigorous comparison of the three estimation methods. As described in detail in Attachment Q, the theoretical comparison also showed that the second and third methods were superior to the first method when a year had 5 or more missing months.

The air quality data base used for this study exhibited the following characteristics: If a station has a bad year, the station has many missing months in that year. If a station has a good (i.e., valid) year, the station seldom has a missing month in that year.

Because of the empirical and theoretical results and the missing data characteristics, we decided to use a combination of the second and third methods, i.e., the second method for an intermediate year, and the third method for either the beginning year or the ending year of each of the two periods, 1966-1972, and 1973-1976.

2.4 Exposure Data Base at Postal Zip-Code Areas

Human exposure to air pollution at each postal zip-code area is estimated by applying a spatial interpolation formula to the exposure values at air monitoring sites. After several trials of applying various spatial interpolation schemes to air quality monitoring data, we decided to use the spatial interpolation program developed by Guardado and Sommers [1977] at the Pacific Southwest Forest and Range Experimental Station, Berkeley, California. The program includes the following:

i) An unevenly spaced orthogonal grid is developed such that each row and column passes through one of the air quality monitoring sites. This grid is called an initial "guess field" grid.

ii) An exposure parameter value at an initial "guess field" grid point is estimated by an inverse square-of-distance weighting interpolation formula applied to the known exposure values at the monitoring sites, i.e.,

$$X = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i}$$

and

$$w_i = \frac{1}{d_i^{\,2}}$$

where X is the interpolated value at the initial "guess field" grid point, X_i are the known exposure values at the monitoring sites, and d_i is the distance between the interpolation point and the ith monitoring site ($i = 1, 2, \ldots, n$).

iii) The initial guess field rows and columns are then adjusted by using the known exposure values and a parabolic leapfrogging technique, which is discussed in Attachment P.

iv) Finally, an exposure value at each postal zip-code area is estimated by a cubic spline interpolation method applied to the corrected field rows and columns.

In actual interpolation, we assigned the smallest measured station value to the extreme northwest, northeast, and southeast corners of the study area. As Figure 1 shows, the study area is bounded by the coastal line in the southeast direction. Therefore, we did not assign any boundary value to the extreme southwest corner. These boundary values were used to force all the interpolated values to be within the range of the measured station values. In addition, the interpolated values at each postal zip-code area are indicated by a confidence level, which was determined by comparing a typical representative distance around a monitor with a distance between the centroid of that zip-code area and the nearest valid monitoring station.

Table 2 presents the typical representative distance around a monitor for each of the six study pollutants. If the distance of the nearest monitoring station is less than or equal to the typical representative distance, the exposure data at that postal zip-code area are designated as "A", the highest confidence level. If the inter-distance is in between the typical representative distance and twice that distance, the exposure data are designated as "B", the moderate confidence level. If the inter-distance exceeds twice the typical representative distance, the data are designated as "C", the lowest confidence level.

Table 2. Typical "Representative Area" of a Monitoring Station Estimated for Each Major Pollutant from the Expected Pollution Gradients of Long-Term Concentration Statistics [After U.S. EPA 1977]

Pollutant	Representative Area Distance from a Station	
OX	10 miles	16 km
NO2	5	8
SO2	3	4.8
CO	1	1.6
TSP	3	4.8
SO4	20	32

The distributions of postal zip-code areas over the study area are plotted in Attachments E and F by using the three confidence levels, A, B, and C. As seen from Attachments E1 and F1 for photochemical oxidants, OX, the great majority of the zip-code areas are classed as A or B, indicating that the exposure values at those zip-code areas are at least believable. On the other hand, Attachments E3 and F3 for carbon monoxide (CO) are filled with the letter "C", indicating that the exposure values are not more than a mere guess. The confidence in the exposure data for other pollutants falls in between those of the above two pollutants (see Attachments E2, E4, F2, F4, F5 and F6).

3. RECOMMENDED DATA BASE IMPROVEMENTS

The human exposure data base developed under the present study will provide researchers of air pollution health effects with a convenient and reliable data source for estimating exposures of individual members of a study population to air pollution from various pollutants. The data base includes monthly exposure data of several parameters for each of the four gaseous (OX, NO2, SO2 and CO) and two particulate (TSP and SO4) pollutants over the entire South Coast Air Basin during the eleven-year period, January 1966 through December 1976.

The exposure parameters include the mean exposure level, the total exposure, and the exceedance frequency and excess exposure whose parameter values are computed for each of four different threshold levels selected for each pollutant. Values of these exposure parameters are given for each of some 400 postal zip-code areas scattered over the study region as well as at each of the air monitoring stations whose air quality data were used to develop data base.

While the present data base is more complete than and superior to those developed by earlier epidemiological studies, the data base needs to be updated periodically and further refined in some exposure estimates:

- By now, the 1977 and 1978 air quality data have become available and the 1979 air quality data will soon become available. These recent air quality data can be used to update the human exposure data base.

- In the development of the present data base, a purely mathematical interpolation scheme was used to estimate exposure values at each postal zip-code area from those computed at air monitoring sites. However, if emissions and meteorological data are also employed in such an estimation process, one may be able to improve exposure estimates, particularly at those postal zip-code areas distant from any air monitoring station.

- A person's work place as well as his residence location can be found through simple survey questionnaires. Then the person's exposure will be better estimated from the sum of the exposure value at his work place and that at his residence than from the latter alone. To facilitate such refinement of exposure estimates, exposure during the typical work hours (say, weekdays 7 A.M. to 5 P.M.) and that during the non-work hours may need to be computed in addition to the present "all day exposure" estimates.

4. REFERENCES

1. Guardado, J.L. and W.T. Sommers, "Interpolation of Unevenly Spaced Data Using a Parabolic Leapfrog Correction Method and Cubic Splines," USDA Forest Service Research Note PSW-324, Pacific Southwest Forest and Range Experiment Station, Berkeley, California, 1977.
2. Horie, Y., J. Cassmassi, L. Lai and L. Gurtowski, "Weekend/Weekday Differences in Oxidants and Their Precursors," EPA-450/4-79-013, U.S. Environmental Protection Agency/OAQPS, Research Triangle Park, North Carolina, March 1979.
3. Lave, L.B. and E.P. Seskin, Air Pollution and Human Health, Johns Hopkins University Press, 1977.
4. Shy, C.M., et al., "The Chattanooga School Children Study: Effects of Community Exposure to Nitrogen Dioxide," J. APCA, Vol. 20, No. 8, August 1970, pp. 539-545.
5. U.S. EPA, "Guideline on Procedures for Constructing Air Pollution Isopleth Profiles and Population Exposure Analysis," EPA-450/2-77-024a (OAQPS No. 1.2-083), U.S. Environmental Protection Agency/OAQPS, Research Triangle Park, North Carolina, October 1977.

Attachments

DESCRIPTION AND SAMPLES OF
HUMAN EXPOSURE DATA BASE

Description and Samples of Human Exposure Data Base

Attachment	Data Base Description
A	Data base documentation
B	Concentration threshold levels used for the human exposure data base
C	Monitoring stations and data availability during 1966-1972
D	Monitoring stations and data availability during 1973-1976
E	Postal zip code area classified according to the distance from the nearest monitoring station during 1966-1972
F	Postal zip code area classified according to the distance from the nearest monitoring station during 1973-1976
	Data Base Samples for Downtown L.A. Site
G	Gaseous Pollutant Percentile Data
H	Particulate Air Quality Data Base
I	Sulfate Air Quality Data
J	Oxidant Exposure Parameter Data Base
K	NO_2 Exposure Parameter Data Base
L	CO Exposure Parameter Data Base
M	SO_2 Exposure Parameter Data Base
N	TSP Exposure Parameter Data Base
O	Sulfate Exposure Parameter Data Base
P	Cubic Spline Interpolation Scheme With A Parabolic Leapfrog Correction
Q	Comparing Methods of Filling In Missing Data

1 HUMAN EXPOSURE DATA BASE

ALL OF THE HUMAN EXPOSURE DATA BASES HAVE THE SAME STYLE HEADER IN THE FORMAT BELOW:

	FIELD	COLUMNS	COMMENTS
1.	VARIABLE	1-3	CO1,NO2,OXI,SO2,TSP,SO4
2.	METHOD CODE	4	A,B,C (SEE METHOD CODE TABLE)
3.	STATION NUMBER OR ZIP	5-9	(SEE STATION NUMBER AND ZIP TABLES)
4.	YEAR	10-11	66-76
5.	MONTH	12-13	1-12

THE EXPOSURE PARAMETER DATA BASE HAS THE FORMAT: (See Attachments J thru M)

1-5.	STANDARD HEADER	1-13	
6.	CONFIDENCE INDICATOR	14-15	1-6 (1=HIGHEST CONFIDENCE)
7.	MEAN	16-20	(F5.1)
8.	DOSAGE (CONC.HRS)	21-26	(I6)
9.	EXCESS FREQUENCY (HRS) > LEVEL 1	27-32	(I6) (SEE EXPOSURE LEVEL TABLE)
10.	EXCESS DOSAGE (CONC.HRS) > LEVEL 1	33-38	(I6)
11.	EXCESS FREQUENCY (HRS) > LEVEL 2	39-44	(I6) (SEE EXPOSURE LEVEL TABLE)
12.	EXCESS DOSAGE (CONC.HRS) > LEVEL 2	45-50	(I6)
13.	EXCESS FREQUENCY (HRS) > LEVEL 3	51-56	(I6) (SEE EXPOSURE LEVEL TABLE)
14.	EXCESS DOSAGE (CONC.HRS) > LEVEL 3	57-62	(I6)
15.	EXCESS FREQUENCY (HRS) > LEVEL 4	63-68	(I6) (SEE EXPOSURE LEVEL TABLE)
16.	EXCESS DOSAGE (CONC.HRS) > LEVEL 4	69-74	(I6)

1 THE PERCENTILE DATA BASE IS IN THE FORMAT: (See Attachment H)

1-5.	STANDARD HEADER	1-13	
6.	CONFIDENCE INDICATOR	14-15	1-6 (1=HIGHEST CONFIDENCE)
7.	MEAN	16-20	(F5.1)
08.	PERCENTILE 1.000	021-023	
09.	PERCENTILE .9975	024-026	
10.	PERCENTILE .9950	027-029	
11.	PERCENTILE .9925	030-032	
12.	PERCENTILE .9900	033-035	
13.	PERCENTILE .9800	036-038	
14.	PERCENTILE .9700	039-041	
15.	PERCENTILE .9600	042-044	
16.	PERCENTILE .9500	045-047	
17.	PERCENTILE .9400	048-050	
18.	PERCENTILE .9300	051-053	
19.	PERCENTILE .9200	054-056	
20.	PERCENTILE .9100	057-059	
21.	PERCENTILE .9000	060-062	
22.	PERCENTILE .8750	063-065	
23.	PERCENTILE .8500	066-068	
24.	PERCENTILE .8250	069-071	

25.	PERCENTILE .8000	072-074
26.	PERCENTILE .7750	075-077
27.	PERCENTILE .7500	078-080
28.	PERCENTILE .7250	081-083
29.	PERCENTILE .7000	084-086
30.	PERCENTILE .6500	087-089
31.	PERCENTILE .6000	090-092
32.	PERCENTILE .5500	093-095
33.	PERCENTILE .5000	096-098
34.	PERCENTILE .4000	099-101
35.	PERCENTILE .3000	102-104
36.	PERCENTILE .2000	105-107
37.	PERCENTILE .1000	108-110
38.	PERCENTILE .0000	111-113

THE S04 AND TSP DATA BASES DO NOT FOLLOW THE PERCENTILE DATA BASE FORMAT BECUASE OF THE LIMITED NUMBER OF MEASUREMENTS (5-6 PER MONTH AVERAGE VERSUS 720 PER MONTH AVERAGE). THE S04 AND TSP DATA BASES HAVE THE FORMAT:

(See Attachments H and I)

1-5.	STANDARD HEADER	1-13		
6.	N - NUMBER OF CASES	14-15	(I2)	(<= 15)
7.	MAX - MAXIMUM VALUE	16-21	(F6.1)	
8.	MIN - MINIMUM VALUE	22-27	(F6.1)	
9.	GMEAN-GEOMETRIC MEAN	28-33	(F6.1)	
10.	G.SD-GEOMETRIC STANDARD DEVIATION	34-39	(F6.1)	
11.	VALUES- ACTUAL MEASUREMENTS	40-130	(15F6.1)	

	Pollutant	Concentration Threshold 1st Level	2nd Level	3rd Level	4th Level
1 hr Concentration	OX	10 pphm (Calif.std.)	15 pphm	20 pphm (1st stage alert)	25 pphm
	NO_2	5 pphm (primary 1 yr. std.)	15 pphm	20 pphm	25 pphm (Calif. 1-Hr. std.)
	SO_2	2 pphm	4 pphm (Calif. 24-Hr. std.)	8 pphm	14 pphm (primary 24-Hr. std.)
	CO	5 ppm	10 ppm (Calif. 12-Hr. std.)	15 ppm	20 ppm
24 hr Concentration	TSP	60 ug/m^3 (Calif. 1 yr. std.)	100 ug/m^3 (Calif. 24-Hr. std.)	150 ug/m^3 (secondary 1 yr. std.)	200 ug/m^3
	SO_4	10 ug/m^3	15 ug/m^3	20 ug/m^3	25 ug/m^3 (Calif. 24-Hr. std.)

Attachment B

Concentration Threshold Levels Used For

Human Exposure Data Base

			66	67	68	69	70	71	72
OXIDANT	30176	ANAHEIM	A	A	A	A	A	A	A
OXIDANT	36151	SAN.BERNARDINO	A	A	A	A	A	A	A
OXIDANT	36165	REDLANDS			A	A	A	A	A
OXIDANT	70001	LOS.ANGELES.DOWNTOWN	A	A	A	A		A	A
OXIDANT	70060	AZUSA	A	A	A	A		A	A
OXIDANT	70069	BURBANK	A	A	A	A		A	A
OXIDANT	70071	WEST.LOS.ANGELES	A	A	A	A		A	A
OXIDANT	70072	LONG.BEACH	A	A	A	A		A	A
OXIDANT	70074	RESEDA	A	A	A	A		A	A
OXIDANT	70075	POMONA	A	A	A	A		A	A
OXIDANT	70076	LENNOX	A	A	A	A		A	A

MEASUREMENT METHOD

OXIDANT A=12. __=Total Oxidant/Colorimetric
B=36. __=Ozone/UV Photometric
C=36.01 =Ozone/Chemiluminescent

Attachment C1. Monitoring Stations Reporting Adequate Amount of Valid Oxidant Data During 1966-1972.

			66	67	68	69	70	71	72
TROGEN DIOXIDE	70001	LOS.ANGELES.DOWNTOWN	A	A	A	A			A
TROGEN DIOXIDE	70060	AZUSA	A	A	A	A		A	A
NITROGEN DIOXIDE	70069	BURBANK	A	A	A	A		A	A
NITROGEN DIOXIDE	70071	WEST.LOS.ANGELES	A	A	A	A		A	A
NITROGEN DIOXIDE	70072	LONG.BEACH	A	A	A	A		A	A
NITROGEN DIOXIDE	70074	RESEDA	A	A	A	A		A	A
NITROGEN DIOXIDE	70075	POMONA	A	A	A	A		A	A
NITROGEN DIOXIDE	70076	LENNOX	A	A	A	A		A	A

MEASUREMENT METHOD

NITROGEN DIOXIDE A=21. =Colorimetric (Saltzman)
B=21.01 =Chemiluminescent

Attachment C2. Monitoring Stations Reporting Adequate Amount of Valid NO_2 Data During 1966-1972.

			66	67	68	69	70	71	72
CARBON MONOXIDE	30176	ANAHEIM	A	A	A	A	A	A	A
CARBON MONOXIDE	36151	SAN.BERNARDINO	A	A	A	A	A	A	A
CARBON MONOXIDE	36165	REDLANDS			A	A	A	A	A
CARBON MONOXIDE	70001	LOS.ANGELES.DOWNTOWN	A	A	A	A		A	A
CARBON MONOXIDE	70060	AZUSA	A	A	A	A		A	A
CARBON MONOXIDE	70069	BURBANK	A	A	A	A		A	A
CARBON MONOXIDE	70071	WEST.LOS.ANGELES	A	A	A	A		A	A
CARBON MONOXIDE	70072	LONG.BEACH	A	A	A	A		A	A
CARBON MONOXIDE	70074	RESEDA	A	A	A	A		A	A
CARBON MONOXIDE	70075	POMONA	A	A	A	A		A	A
CARBON MONOXIDE	70076	LENNOX	A	A	A	A		A	A

MEASUREMENT METHOD

Carbon Monoxide A=15. =Long Path NDIR
B=15.0T =FID

Attachment C3. Monitoring Stations Reporting Adequate Amount of Valid CO Data During 1966-72.

			66	67	68	69	70	71	72
SULFUR DIOXIDE	30176	ANAHEIM	A	A	A	A	A	A	A
SULFUR DIOXIDE	70001	LOS ANGELES DOWNTOWN	A	A	A	A		A	A

MEASUREMENT METHOD

SULFUR DIOXIDE A=18._=COULOMETRIC
B=18.$\overline{01}$=FLAME PHOTOMETRIC

Attachment C4. Monitoring Stations Reporting Adequate Amount of Valid SO_2 Data During 1966-1972.

			73	74	75	76
OXIDANT	30176	ANAHEIM	A	A		A
OXIDANT	30177	LA.HABRA	A	A	A	A
OXIDANT	30185	COSTA.MESA.HARBOR	A	A		A
OXIDANT	30186	EL.TORO		B	B	B
OXIDANT	30190	LOS.ALAMITOS.ORANGEWOOD		B	B	B
OXIDANT	33139	INDIO.OASIS.ST	A	A	A	
OXIDANT	33141	HEMET.STATE.ST	B	B	B	B
OXIDANT	33144	RIVERSIDE.RUBIDOUX	A	A	A	
OXIDANT	33146	RIVERSIDE.MAGNOLIA	A	A	A	
OXIDANT	33149	PERRIS		B	B	B
OXIDANT	36151	SAN.BERNARDINO		C	C	C
OXIDANT	36165	REDLANDS	C	C	C	C
OXIDANT	36173	CHINO.RIVERSIDE.AVE	C	C	C	C
OXIDANT	36174	UPLAND.CIVIC.CENTER	C	C	C	
OXIDANT	36175	UPLAND.ARB	A	A	A	
OXIDANT	36176	FONTANA.FOOTHILL		C	C	C
OXIDANT	56402	OJAI	C	C	C	
OXIDANT	56404	SANTA.PAULA	C	C	C	
OXIDANT	70001	LOS.ANGELES.DOWNTOWN	A	A	A	A
OXIDANT	70060	AZUSA	A	A	A	A
OXIDANT	70069	BURBANK	A	A	A	A
OXIDANT	70071	WEST.LOS.ANGELES	A	A	A	A
OXIDANT	70072	LONG.BEACH	A	A	A	A
OXIDANT	70074	RESEDA	A	A	A	A
OXIDANT	70075	POMONA	A	A	A	A
OXIDANT	70076	LENNOX	A	A	A	A
OXIDANT	70080	WHITTIER	A	A	A	A
OXIDANT	70081	NEWHALL	A	A	A	A
OXIDANT	70083	PASADENA.WALNUT	A	A	A	A
OXIDANT	70084	LYNWOOD		A	A	A

MEASUREMENT METHOD

OXIDANT

A=12.__=TOTAL OXIDANT/COLORIMETRIC
B=36.__=OZONE/UV PHOTOMETRIC
C=36.01=OZONE/CHEMILUMINESCENT

Attachment D1. Monitoring Stations Reporting Adequate Amount of Valid Oxidant Data During 1973-1976.

			73	74	75	76
NITROGEN.DIOXIDE	30176	ANAHEIM	A	A	A	A
NITROGEN.DIOXIDE	30177	LA.HABRA	A	A	A	A
TROGEN.DIOXIDE	30185	COSTA.MESA.HARBOR	A	A	A	A
NITROGEN.DIOXIDE	33139	INDIO.OASIS.ST	A	A	A	
NITROGEN.DIOXIDE	33144	RIVERSIDE.RUBIDOUX	A	A	A	A
NITROGEN.DIOXIDE	33146	RIVERSIDE.MAGNOLIA	A	A	A	
NITROGEN.DIOXIDE	36151	SAN.BERNARDINO	A	A	A	A
NITROGEN.DIOXIDE	36165	REDLANDS	A	A		A
NITROGEN.DIOXIDE	70001	LOS.ANGELES.DOWNTOWN	A	A	A	A
NITROGEN.DIOXIDE	70060	AZUSA	A	A	A	A
NITROGEN.DIOXIDE	70069	BURBANK	A	A	A	A
NITROGEN.DIOXIDE	70071	WEST.LOS.ANGELES	A	A	A	A
NITROGEN.DIOXIDE	70072	LONG.BEACH	A	A	A	A
NITROGEN.DIOXIDE	70074	RESEDA	A	A	A	A
NITROGEN.DIOXIDE	70075	POMONA	A	A	A	A
NITROGEN.DIOXIDE	70076	LENNOX	A	A	A	A
NITROGEN.DIOXIDE	70080	WHITTIER	A	A	A	A
NITROGEN.DIOXIDE	70081	NEWHALL	A	A	A	A
NITROGEN.DIOXIDE	70083	PASADENA.WALNUT	A	A	A	A
NITROGEN.DIOXIDE	70084	LYNWOOD		A	A	A

MEASUREMENT METHOD

NITROGEN DIOXIDE A=21.__=COLORIMETRIC(SALTZMAN)
B=21.01=CHEMILUMINESCENT

Attachment D2. Monitoring Stations Reporting Adequate Amount of Valid NO_2 Data During 1973-1976.

			73	74	75	76
CARBON.MONOXIDE	30176	ANAHEIM	A	A	A	A
CARBON.MONOXIDE	30177	LA.HABRA	A	A		A
CARBON.MONOXIDE	30185	COSTA.MESA.HARBOR	A	A		A
CARBON.MONOXIDE	30189	LAGUNA.BEACH.BROADWAY		A	A	A
CARBON.MONOXIDE	33139	INDIO.OASIS.ST	A	A	A	
CARBON.MONOXIDE	33140	NORCO.PRADO.PARK		A	A	A
CARBON.MONOXIDE	33144	RIVERSIDE.RUBIDOUX	A	A	A	A
CARBON.MONOXIDE	33146	RIVERSIDE.MAGNOLIA		A	A	A
CARBON.MONOXIDE	36151	SAN.BERNARDINO	A	A	A	A
CARBON.MONOXIDE	36165	REDLANDS	A		A	A
CARBON.MONOXIDE	36173	CHINO.RIVERSIDE.AVE	A	A	A	A
CARBON.MONOXIDE	36174	UPLAND.CIVIC.CENTER	A	A	A	
CARBON.MONOXIDE	36175	UPLAND.ARB	A	A	A	
CARBON.MONOXIDE	36176	FONTANA.FOOTHILL		A	A	A
CARBON.MONOXIDE	70001	LOS.ANGELES.DOWNTOWN	A	A	A	A
CARBON.MONOXIDE	70060	AZUSA	A	A	A	A
CARBON.MONOXIDE	70069	BURBANK	A	A	A	A
CARBON.MONOXIDE	70071	WEST.LOS.ANGELES	A	A	A	A
CARBON.MONOXIDE	70072	LONG.BEACH	A	A	A	A
CARBON.MONOXIDE	70074	RESEDA	A	A	A	A
CARBON.MONOXIDE	70075	POMONA	A	A	A	A
CARBON.MONOXIDE	70076	LENNOX	A	A	A	A
CARBON.MONOXIDE	70080	WHITTIER	A	A	A	A
CARBON.MONOXIDE	70081	NEWHALL	A	A	A	A
CARBON.MONOXIDE	70083	PASADENA.WALNUT	A	A	A	A
CARBON.MONOXIDE	70084	LYNWOOD		A	A	A

MEASUREMENT METHOD

CARBON MONOXIDE A=15.__=LONG PATH NDIR
B=15.01=FID

Attachment D3. Monitoring Stations Reporting Adequate Amount of Valid CO Data During 1973-1976.

			73	74	75	76
SULFUR.DIOXIDE	30176	ANAHEIM	A	A	A	A
SULFUR.DIOXIDE	30177	LA.HABRA	A	A	A	A
SULFUR.DIOXIDE	30185	COSTA.MESA.HARBOR	A	A		A
SULFUR.DIOXIDE	30190	LOS.ALAMITOS.ORANGEWOOD		A	A	A
SULFUR.DIOXIDE	36151	SAN.BERNARDINO	A	A	A	A
SULFUR.DIOXIDE	70001	LOS.ANGELES.DOWNTOWN	A	A	A	A
SULFUR.DIOXIDE	70060	AZUSA	A	A	A	A
SULFUR.DIOXIDE	70069	BURBANK	A	A	A	A
SULFUR.DIOXIDE	70071	WEST.LOS.ANGELES	A	A	A	A
SULFUR.DIOXIDE	70072	LONG.BEACH	A	A	A	A
SULFUR.DIOXIDE	70074	RESEDA	A	A	A	A
SULFUR.DIOXIDE	70075	POMONA	A	A	A	A
SULFUR.DIOXIDE	70076	LENNOX	A	A	A	A
SULFUR.DIOXIDE	70080	WHITTIER	A	A	A	A
SULFUR.DIOXIDE	70081	NEWHALL	A	A	A	A
SULFUR.DIOXIDE	70083	PASADENA.WALNUT	A	A	A	A
SULFUR.DIOXIDE	70084	LYNWOOD		A	A	A

MEASUREMENT METHOD

SULFUR DIOXIDE A=18.__=COULOMETRIC
B=18.01=FLAME PHOTOMETRIC

Attachment D4. Monitoring Stations Reporting Adequate Amount of Valid SO_2 Data During 1973-1976.

			73	74	75	76
TSP	30176	ANAHEIM	A	A	A	A
TSP	30177	LA HABRA	A	A	A	A
TSP	30184	LOS ALAMITOS AIR STATION		A	A	A
TSP	30185	COSTA MESA HARBOR	A	A	A	A
TSP	30186	EL TORO	A	A	A	A
TSP	30189	LAGUNA BEACH BROADWAY		A	A	A
TSP	30190	LOS ALAMITOS ORANGEWOOD		A	A	A
TSP	33126	RIVERSIDE 11th ST MALL		A	A	
TSP	33142	RIVERSIDE TRAILER		A	A	A
TSP	33144	RIVERSIDE RUBIDOUX	A	A	A	A
TSP	33146	RIVERSIDE MAGNOLIA		A	A	A
TSP	36151	SAN BERNARDINO	A	A	A	A
TSP	56401	VENTURA	A	A	A	A
TSP	56402	OJAI	A	A	A	A
TSP	56404	SANTA PAULA	A	A	A	A
TSP	56408	CAMARILLO PALM	A	A		A
TSP	56409	POINT MUGU	A	A	A	A
TSP	56410	OXNARD A ST	A	A	A	A
TSP	56412	PORT HUENEME	A	A	A	A
TSP	56413	SIMI VALLEY	A	A	A	A
TSP	56414	VENTURA TELEGRAPH RD	A	A	A	A
TSP	56415	THOUSAND OAKS WINDSOR	A	A	A	A
TSP	70001	LOS ANGELES DOWNTOWN	A	A	A	A
TSP	70060	AZUSA	A	A	A	A
TSP	70076	LENNOX	A	A	A	A
TSP	70083	PASADENA WALNUT	A	A	A	A
			A	A	A	A

MEASUREMENT METHOD

TSP A = HIGH VOLUME SAMPLING

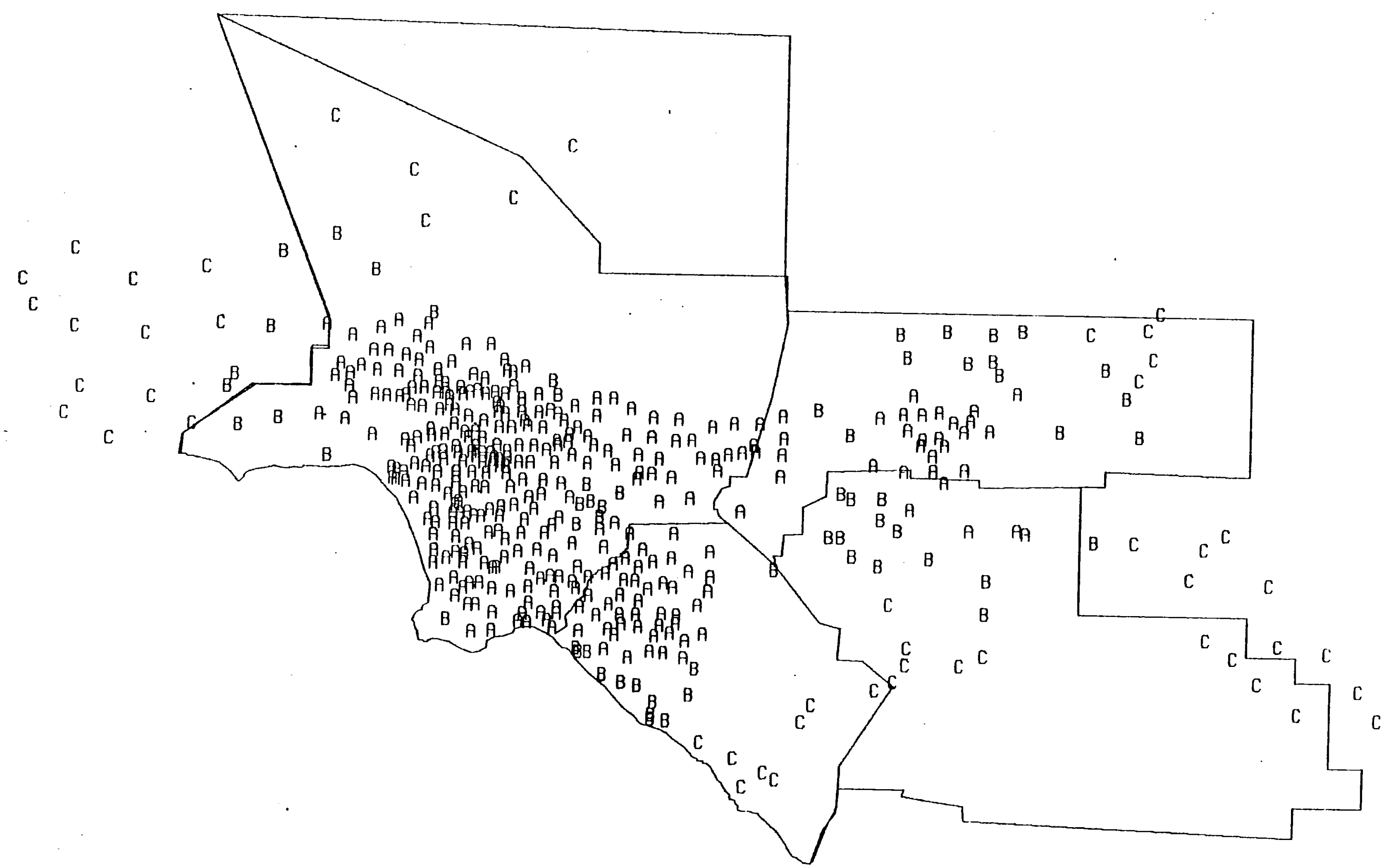
Attachment D5. Monitoring Stations Reporting Adequate Amount of Valid TSP Data During 1973-1976.

			73	74	75	76
S04	70001	LOS ANGELES DOWNTOWN	A	A	A	A
S04	70060	AZUSA	A	A	A	A
S04	70071	WEST LOS ANGELES	A	A	A	A
S04	70074	RESEDA	A	A	A	A
S04	70076	LENNOX	A	A	A	A
S04	70083	PASADENA WALNUT	A	A	A	A
S04	70084	LYNWOOD	A	A	A	A

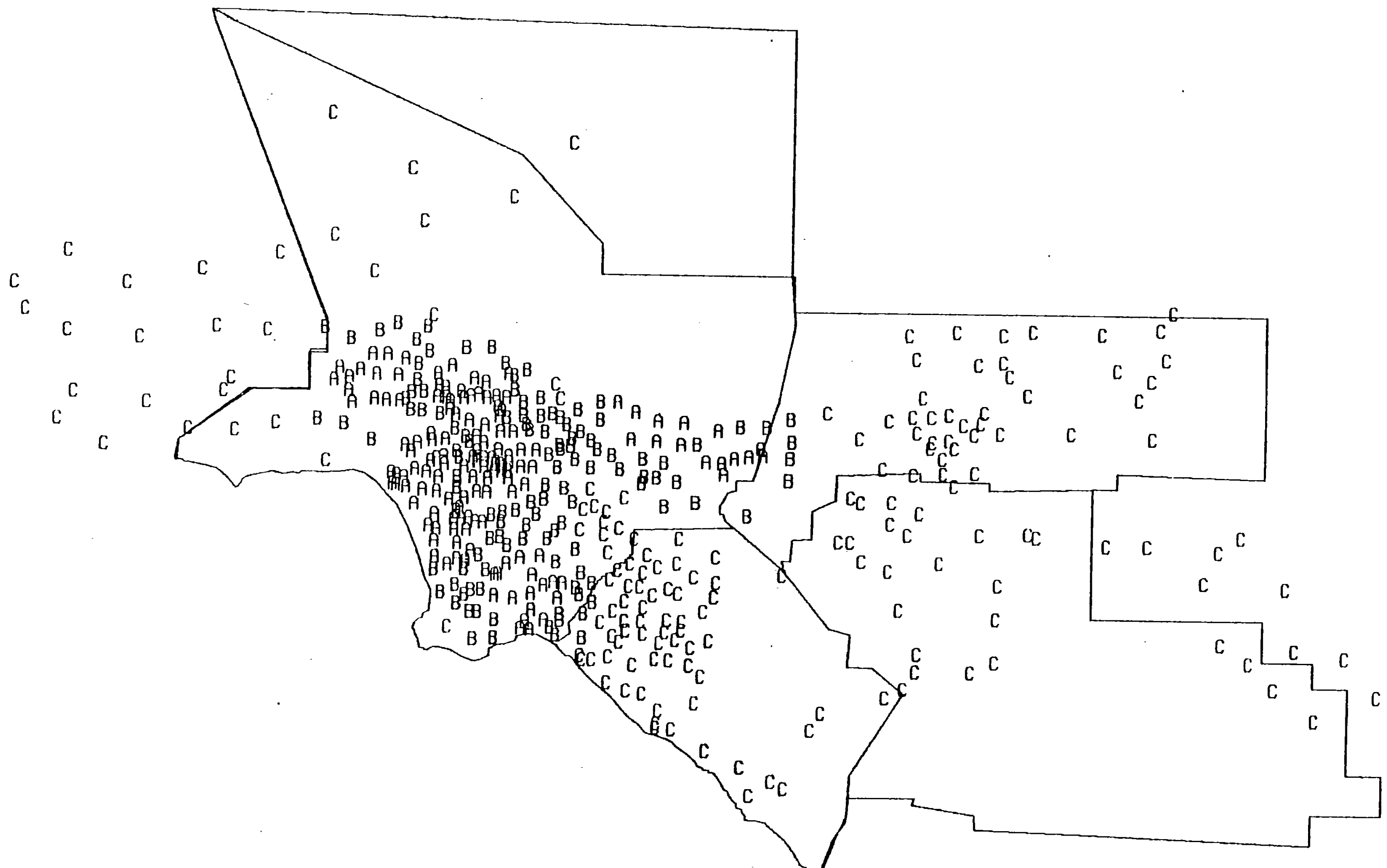
MEASUREMENT METHOD

SULFATE A = HIGH VOLUME SAMPLING

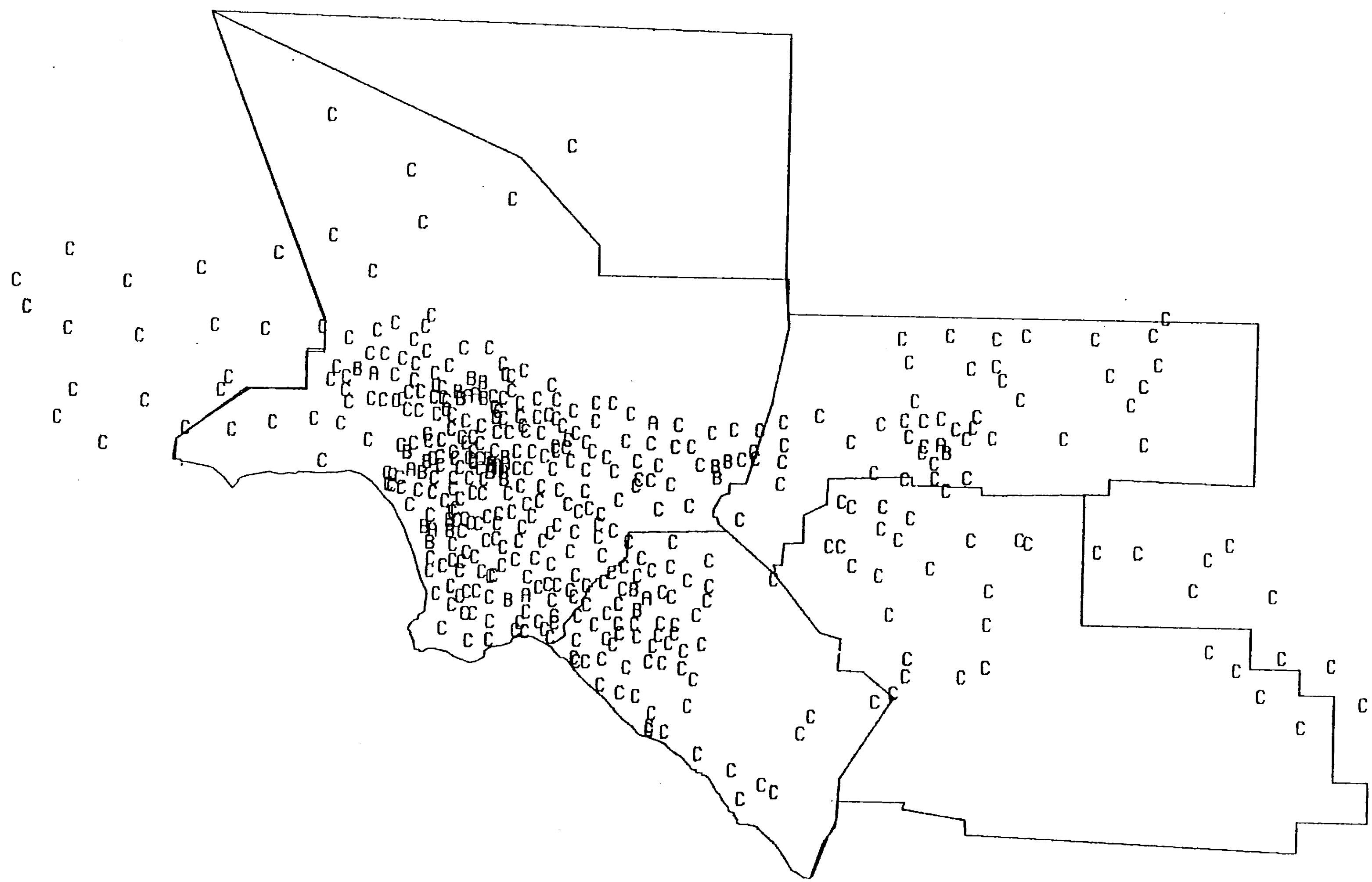
Attachment D6. Monitoring Stations Reporting Adequate Amount of Valid Sulfate Data During 1973-1976.



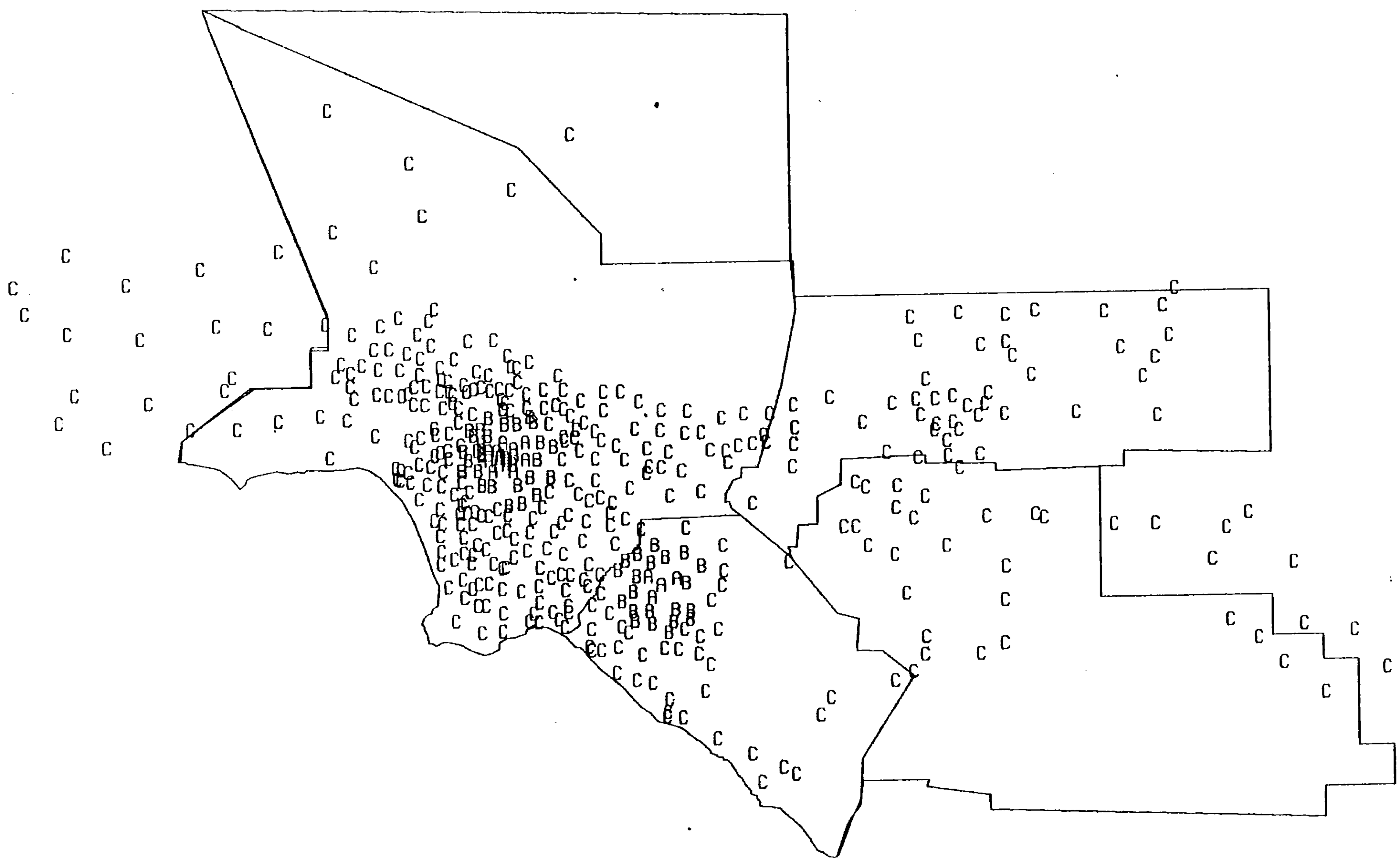
Attachment E1 Postal Zip Code Area Classified According to the Distance From the Nearest Air Monitoring Station for Oxidant during 1966-72 ($A \leq 16$ Km, 16 Km $< B \leq 32$ Km, $C > 32$ Km).



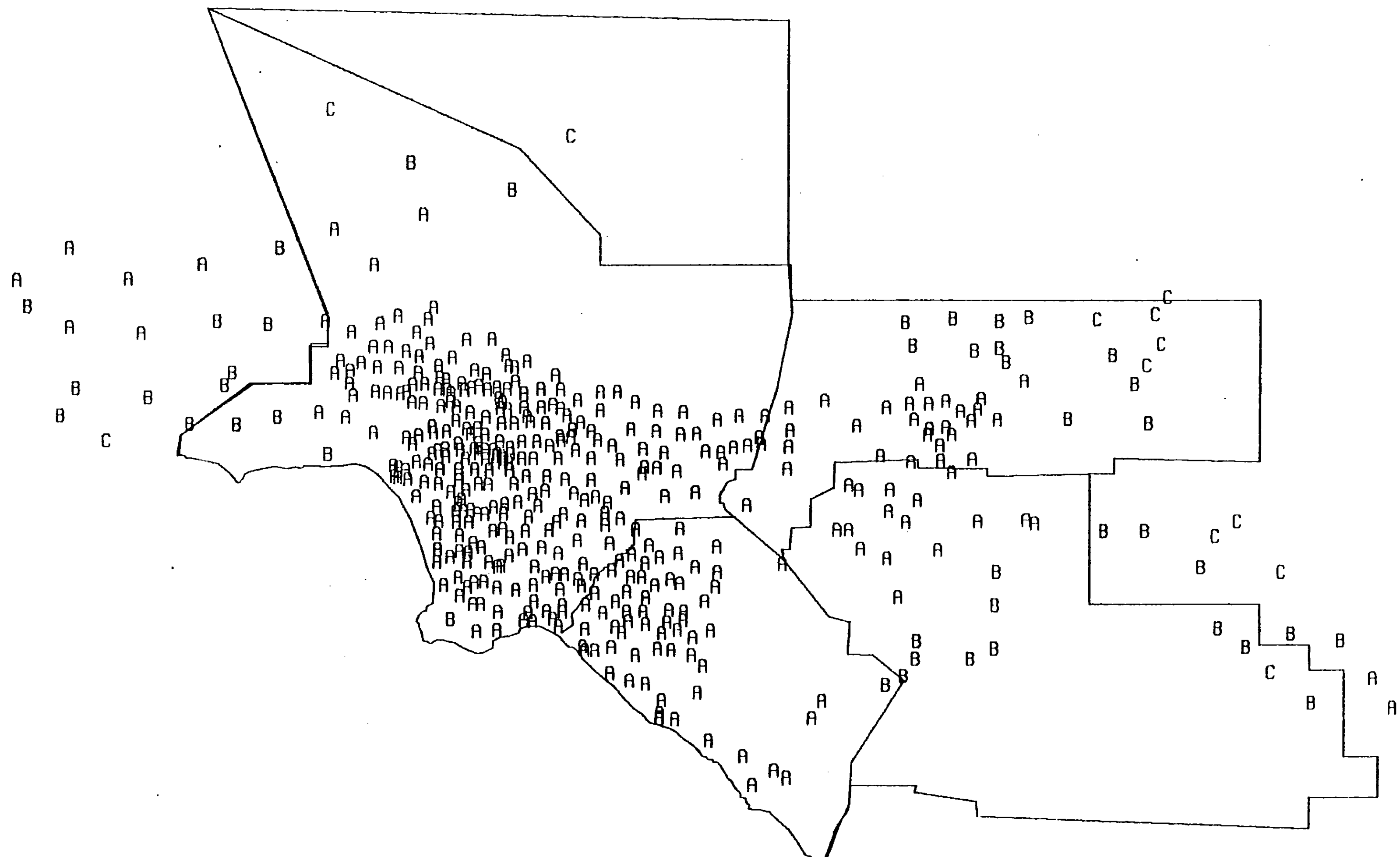
Attachment E2. Postal Zip Code Area Classified According to the Distance from the Nearest Air Monitoring Station for NO_2 during 1966-72 ($A \leq 8$ Km, 8 Km $< B \leq 16$ Km, $C > 16$ Km).



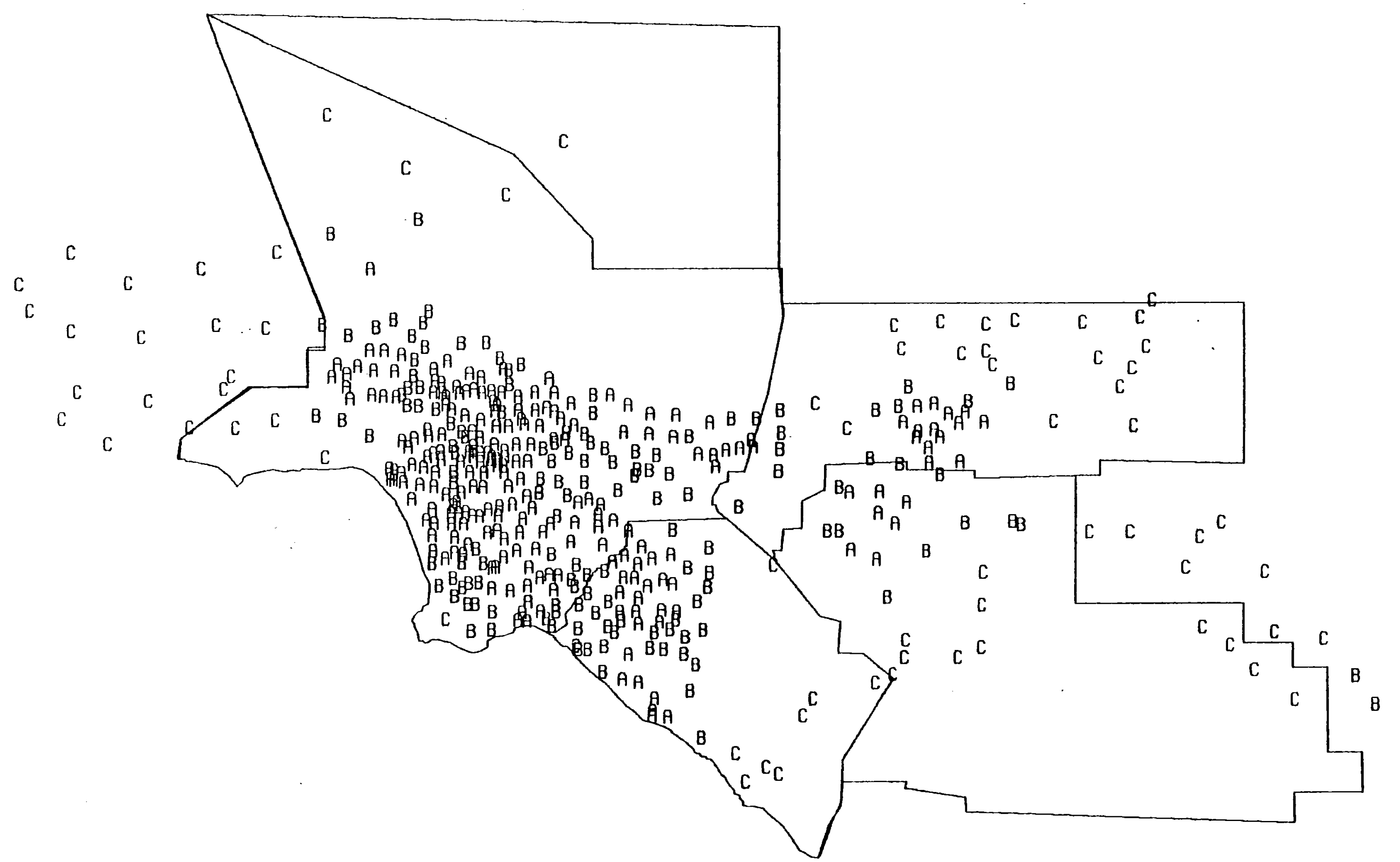
Attachment E3 Postal Zip Code Area Classified According to the Distance From the Nearest Air Monitoring Station for CO during 1966-72 (A ≤ 1.6 Km, 1.6 Km < B ≤ 3.2 Km, C > 3.2 Km).



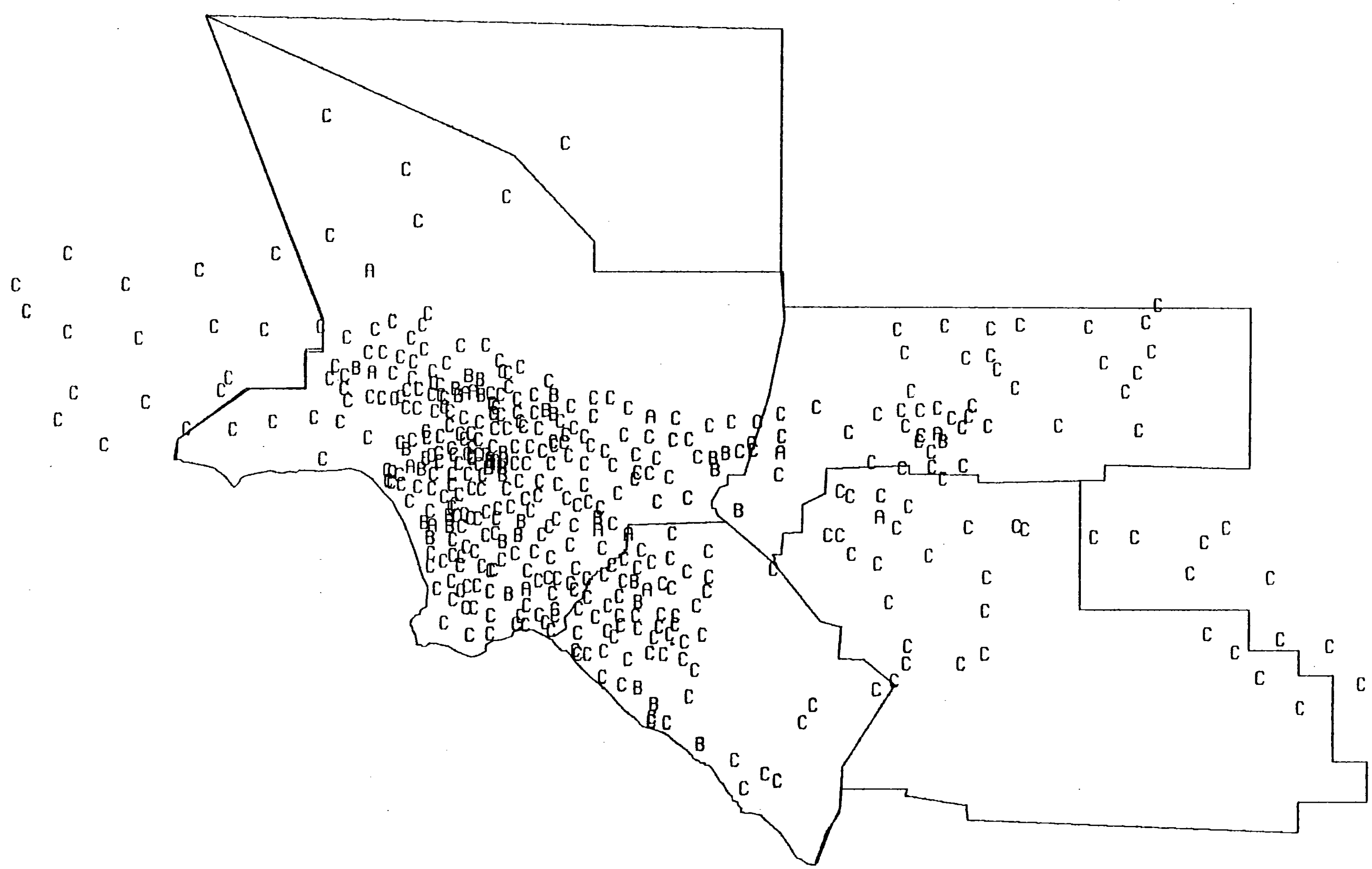
Attachment E4 Postal Zip Code Area Classified According to the Distance from the Nearest Air Monitoring Station for SO_2 during 1966-72 (A $\leq$ 4.8 Km, 4.8 Km < B $\leq$ 9.6 Km, C > 9.6 Km).



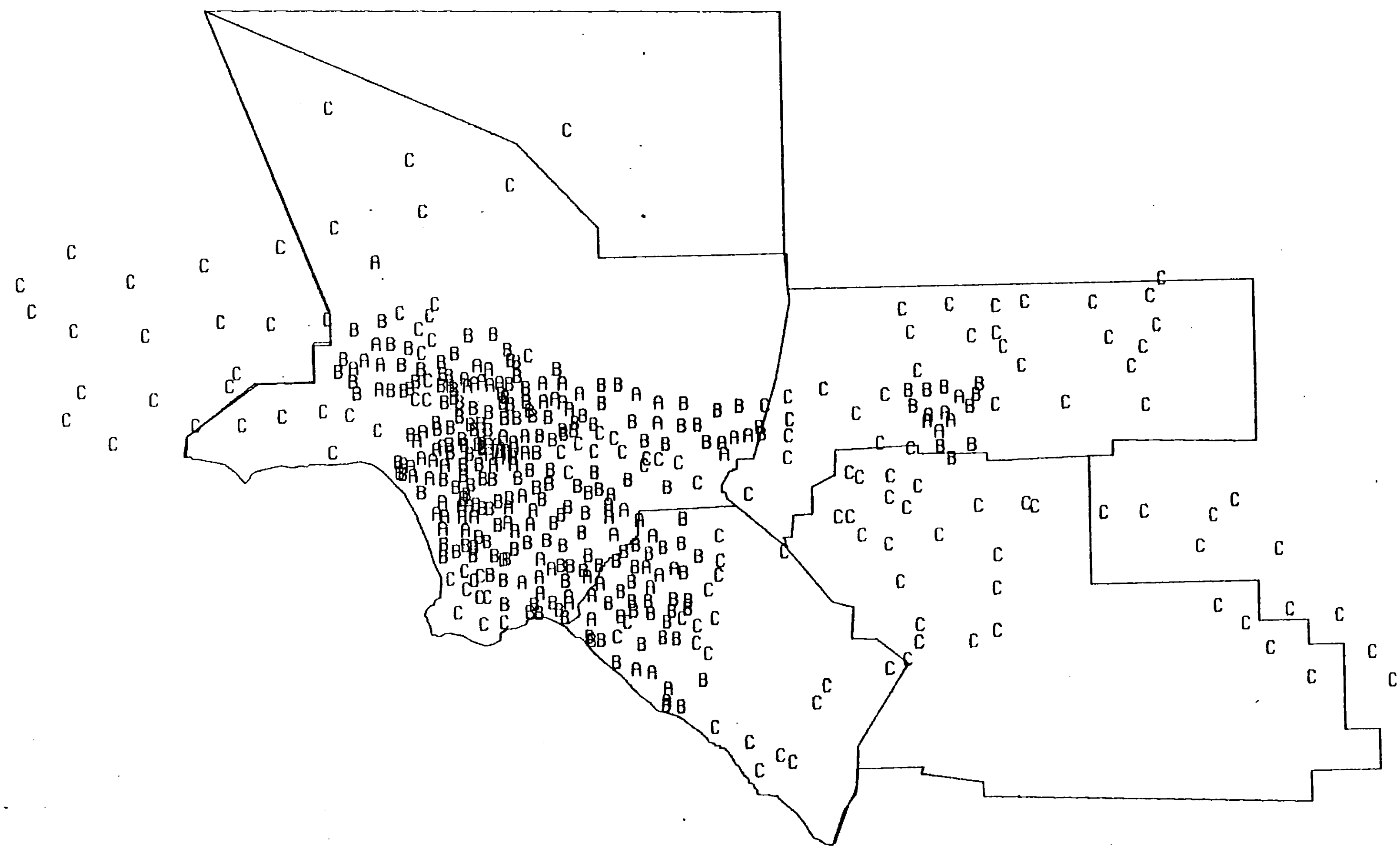
Attachment F1 Postal Zip Code Area Classified According to the Distance from the Nearest Air Monitoring Station for Oxidant during 1973-76 (A ≤ 16 Km,32 Km < B ≤ 32 Km,C > 32 Km).



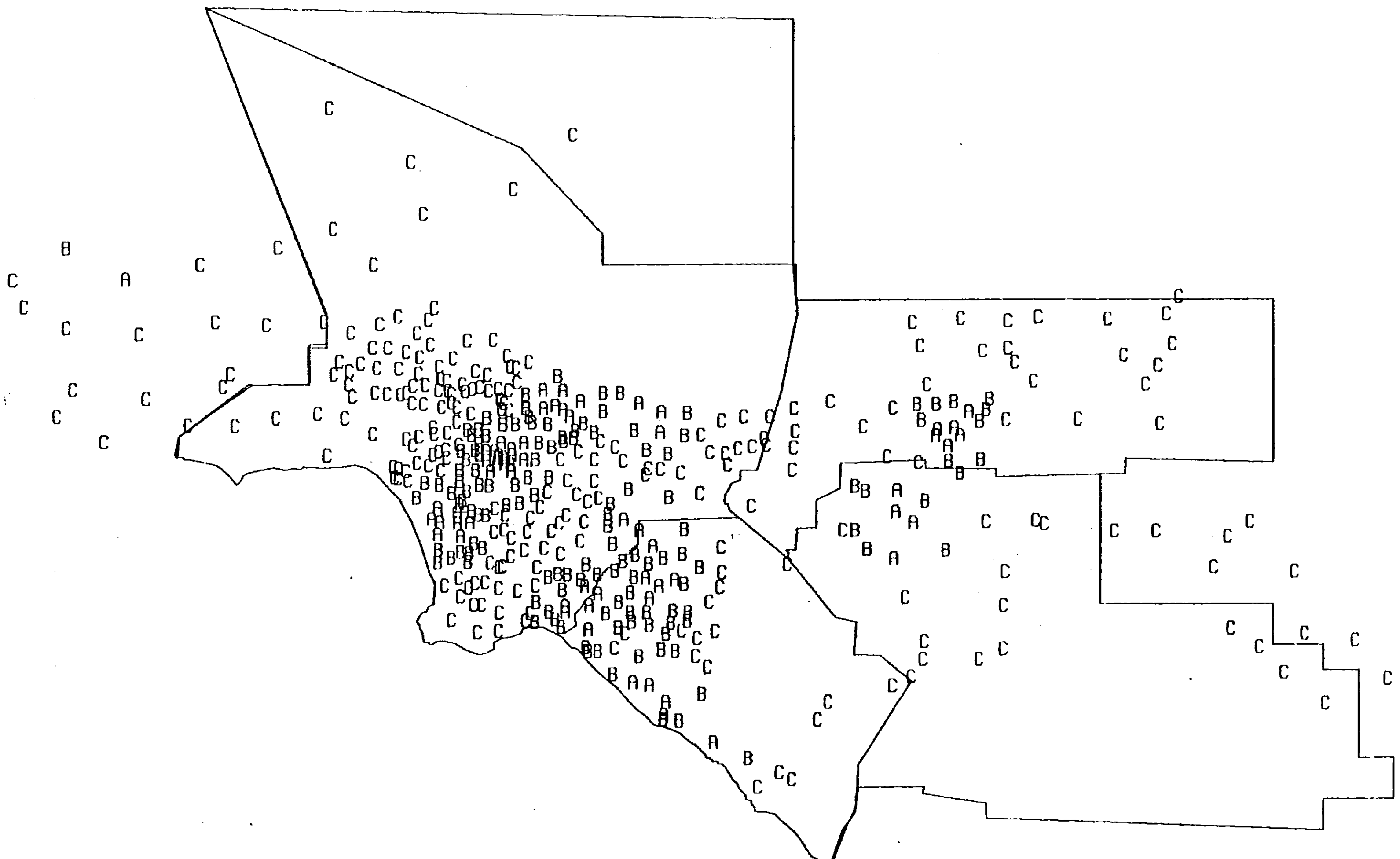
Attachment F2 Postal Zip Code Area Classified According to the Distance from the Nearest Air Monitoring Station for NO_2 during 1973-76 ($A \leq 8$ Km, 8 Km $< B \leq 16$ Km, $C > 16$ Km).



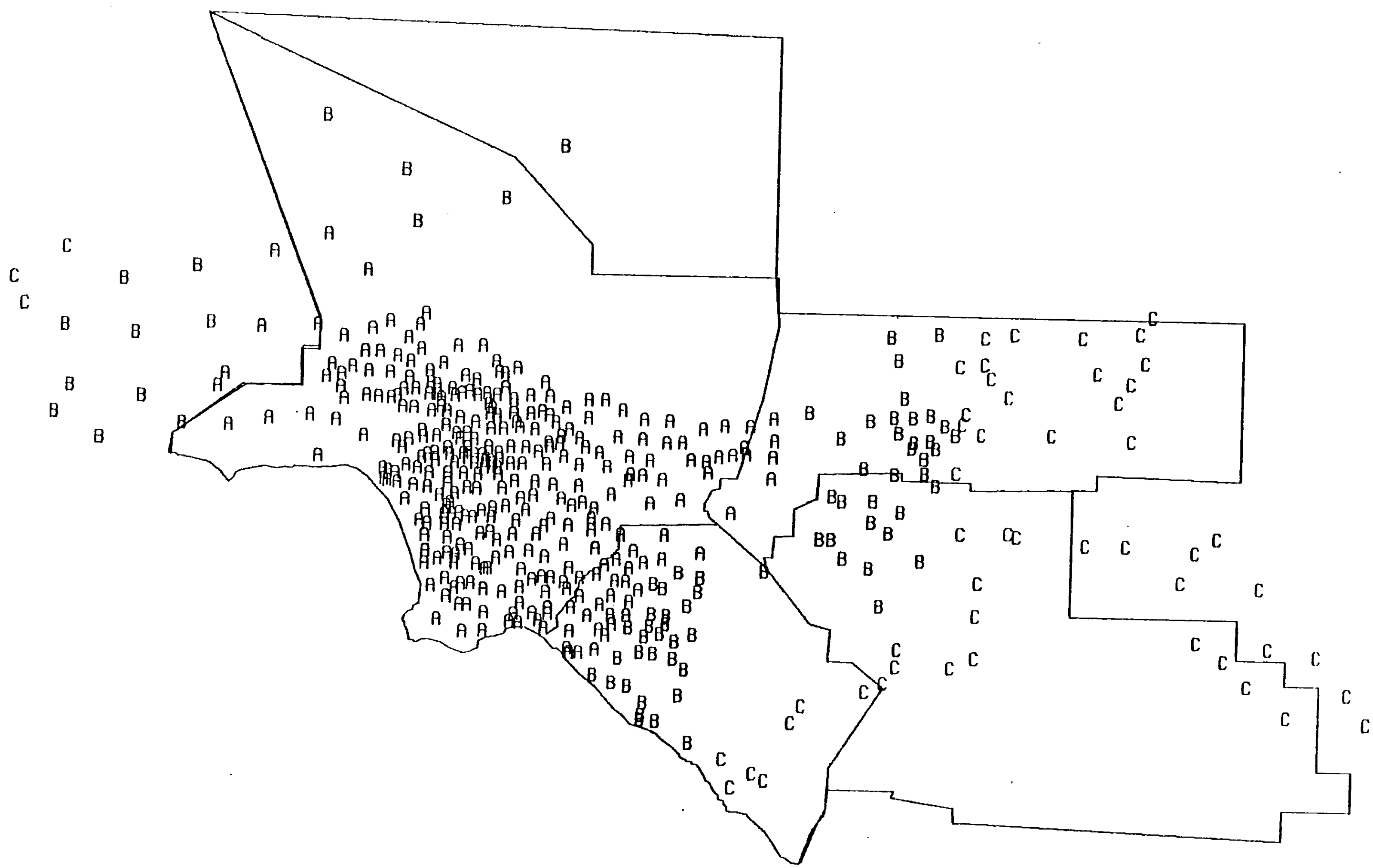
Attachment F3 Postal Zip Code Area Classified According to the Distance from the Nearest Air Monitoring Station for CO during 1973-76 (A ≤ 1.6 Km, 1.6 Km < B ≤ 3.2 Km, C > 3.2 Km).



Attachment F4. Postal Zip Code Area Classified According to the Distance from the Nearest Air Monitoring Station for SO_2 during 1973-76 ($A \leq 4.8$ Km, 4.8 Km $< B \leq 9.6$ Km, $C > 9.6$ Km).



Attachment F5 Postal Zip Code Area Classified According to the Distance from the Nearest Air Monitoring Station for TSP during 1973-76 ($A \leq 4.8$ Km, $4.8 \text{ Km} < B \leq 9.6$ Km, $C > 9.6$ Km).



Attachment F6 Postal Zip Code Area Classified According to the Distance of the Nearest Air Monitoring Station for SO_4 during 1973-76 ($A \leq 32$ Km, $32 \text{ Km} < B \leq 64$ Km, $C > 64$ Km).

G

			MEAN	1.000	.9975	.9950	.9925	.9900	.9800	.9700	.9600	.9500	.9400	.9300	.9200	.9100	.9000	.8750	.8500	.8250	.8000	.7750	.7500	.7250	.7000	.6500	.6000	.5500	.5000	.4000	.3000	.2000	.1000	.0000
				PERCENTILE VALUES																														
C01A7000166	1	1	12.2	36	32	32	30	27	25	24	23	22	22	21	20	20	19	18	18	17	16	16	15	14	14	13	12	12	11	10	9	8	7	4
C01A7000167	1	1	12.3	40	38	30	29	26	25	24	23	23	22	21	20	20	19	19	18	17	16	15	15	14	14	13	12	11	11	10	9	8	7	6
C01A7000168	1	1	11.9	32	29	25	24	23	22	21	20	20	20	19	19	19	18	17	16	15	15	14	14	13	13	12	12	11	11	10	9	9	8	7
C01A7000169	1	1	7.4	36	32	28	23	23	21	19	17	17	16	16	15	14	14	13	12	12	11	10	10	9	9	8	7	7	6	5	4	3	3	2
C01A7000170	1	3	7.3	33	31	26	23	23	21	19	17	17	16	16	15	14	14	13	12	12	11	10	10	9	9	8	7	7	6	5	4	3	3	2
C01A7000171	1	1	7.3	30	29	24	23	23	21	19	18	17	16	15	14	14	13	13	12	11	10	10	9	9	8	8	7	6	6	5	5	4	3	1
C01A7000172	1	3	6.4	27	26	23	22	20	18	17	16	15	14	13	13	12	12	11	11	10	9	9	8	8	7	7	6	6	6	5	5	4	3	1
C01A7000173	1	1	5.6	24	23	22	20	17	15	14	14	13	12	11	11	10	10	9	9	8	8	7	7	6	6	6	5	5	5	4	4	3	3	1
C01A7000174	1	1	6.4	22	21	20	19	19	18	17	15	13	13	12	12	12	11	11	10	10	9	9	8	8	8	7	7	6	6	5	4	3	2	1
C01A7000175	1	1	8.4	40	32	28	28	25	23	21	20	19	18	17	17	16	15	15	14	13	12	11	11	10	10	9	8	8	7	6	5	4	3	1
C01A7000176	1	1	6.7	22	22	21	20	19	17	15	15	14	14	13	13	12	12	11	11	10	10	9	9	8	8	7	7	6	6	5	4	3	3	1
C01A7000166	2	1	8.9	28	28	23	21	21	18	18	17	16	16	16	15	15	14	13	13	12	12	11	11	11	10	10	9	8	8	7	7	6	5	3
C01A7000167	2	1	11.3	30	28	26	26	25	23	21	20	20	19	19	18	17	17	16	15	14	14	14	13	13	13	12	11	11	10	9	9	8	7	6
C01A7000168	2	1	11.6	27	25	25	24	24	21	20	20	19	18	17	17	16	16	15	15	15	14	14	13	13	13	12	12	11	11	10	9	9	8	7
C01A7000169	2	1	6.8	24	22	20	20	19	17	15	14	13	13	13	12	12	11	10	9	9	8	8	8	8	7	7	7	6	6	5	5	4	4	1
C01A7000170	2	3	6.2	24	22	21	20	19	17	15	14	13	13	13	12	12	11	10	9	9	8	8	8	7	7	7	6	6	6	5	4	4	3	1
C01A7000171	2	1	5.6	25	22	22	20	19	16	15	14	13	13	12	12	11	11	10	9	8	8	7	7	6	6	6	5	5	5	4	3	3	2	1
C01A7000172	2	3	5.2	21	20	19	17	17	14	14	13	12	12	11	11	10	10	9	9	8	8	7	7	6	6	6	5	5	5	4	3	3	2	1
C01A7000173	2	1	4.9	17	17	15	14	14	12	12	11	10	10	9	9	9	9	8	8	7	7	6	6	6	6	5	5	5	4	4	3	3	2	1
C01A7000174	2	1	4.9	20	18	18	17	17	15	13	12	11	11	10	10	9	9	9	8	7	7	7	6	6	6	5	5	4	4	3	3	2	2	1
C01A7000175	2	1	6.0	26	24	22	22	20	16	16	14	14	13	12	12	12	11	11	10	9	8	8	8	7	7	6	6	5	5	4	4	3	2	1
C01A7000176	2	1	4.0	21	18	16	15	14	12	11	10	9	9	9	9	8	8	7	7	6	6	6	5	5	5	4	4	4	3	3	2	2	1	0
C01A7000166	3	1	10.0	25	23	23	22	21	19	18	17	16	16	16	15	15	15	14	13	13	12	12	11	11	11	10	10	9	9	8	8	7	7	4
C01A7000167	3	1	9.1	25	22	21	20	20	17	16	16	15	15	14	14	14	13	13	13	12	11	11	11	10	10	9	9	9	8	8	7	7	6	3
C01A7000168	3	1	9.3	22	20	20	19	18	17	16	15	14	14	14	13	13	13	12	12	11	11	11	10	10	10	9	9	9	9	8	8	7	7	6
C01A7000169	3	1	6.6	19	17	17	16	16	14	13	13	12	12	12	11	11	11	10	10	9	9	8	8	8	8	7	7	6	6	5	5	4	3	2
C01A7000170	3	3	5.6	20	18	17	16	16	13	12	12	11	11	10	10	10	10	9	9	8	8	7	7	7	7	6	6	5	5	5	4	4	3	2
C01A7000171	3	1	4.7	21	19	17	17	15	12	10	10	9	9	8	8	8	8	7	7	6	6	6	6	5	5	5	5	4	4	4	3	3	2	1
C01A7000172	3	3	4.2	17	16	14	14	13	11	9	9	9	8	8	8	8	7	7	6	6	6	6	6	5	5	5	4	4	4	4	3	3	2	1
C01A7000173	3	1	3.7	12	12	10	10	10	9	8	8	8	7	7	7	7	6	6	5	5	5	5	5	4	4	4	3	3	3	3	3	2	2	1
C01A7000174	3	1	3.3	20	16	15	13	12	10	10	8	8	7	7	7	6	6	6	5	5	4	4	4	4	4	4	3	3	3	2	2	2	1	1
C01A7000175	3	1	4.5	20	19	18	16	15	12	11	10	10	10	9	9	9	8	8	7	7	6	6	6	5	5	5	4	4	4	3	3	2	2	1
C01A7000176	3	1	3.8	14	13	12	11	11	10	9	9	8	8	8	7	7	7	6	6	6	5	5	5	5	5	4	4	4	3	3	2	2	1	0
C01A7000166	4	1	9.1	21	20	18	18	18	17	16	16	15	14	14	13	13	13	12	12	11	11	10	10	10	9	9	9	8	8	8	8	7	7	5
C01A7000167	4	1	8.0	15	14	13	13	13	12	12	12	12	11	11	11	10	10	10	9	9	9	9	9	9	9	8	8	8	8	7	7	7	6	5
C01A7000168	4	1	4.1	18	15	15	14	14	13	12	11	10	10	10	9	9	9	8	8	8	7	7	6	6	5	5	4	3	3	2	1	1	1	1
C01A7000169	4	1	4.7	17	13	13	12	12	12	10	9	9	9	9	8	8	8	7	7	6	6	6	6	6	5	5	4	4	4	3	3	3	2	1
C01A7000170	4	3	4.5	17	14	13	12	12	12	10	9	9	9	9	8	8	8	7	7	6	6	6	6	6	5	5	5	4	4	4	3	3	2	1
C01A7000171	4	1	4.3	17	16	13	13	12	11	11	9	9	9	8	8	8	8	7	7	7	6	6	6	5	5	5	4	4	4	3	3	2	2	1
C01A7000172	4	3	4.2	16	14	13	12	12	11	10	9	9	9	8	8	8	8	7	7	7	6	6	6	5	5	5	4	4	4	3	3	2	2	1
C01A7000173	4	1	4.1	15	12	12	11	11	10	9	9	8	8	8	7	7	7	6	6	6	5	5	5	5	5	4	4	4	4	3	3	3	2	1
C01A7000174	4	1	2.9	11	11	10	9	9	7	7	6	6	6	6	5	5	5	5	5	4	4	4	4	4	3	3	3	3	3	2	2	1	1	1
C01A7000175	4	1	3.2	17	16	16	13	12	9	9	8	7	7	7	7	6	6	6	5	5	5	4	4	4	4	4	3	3	3	2	2	1	1	1
C01A7000176	4	1	3.3	14	13	11	10	9	8	8	7	7	6	6	6	6	6	5	5	5	5	4	4	4	4	4	3	3	3	3	2	2	1	0
C01A7000166	5	1	8.2	19	17	16	16	15	13	12	12	12	11	11	11	11	10	10	10	9	9	9	9	9	9	8	8	8	8	8	7	7	6	4
C01A7000167	5	1	8.5	18	17	17	16	16	15	14	13	12	12	12	12	11	11	11	11	10	10	10	9	9	9	9	9	8	8	8	7	7	6	4
C01A7000168	5	1	3.3	15	13	13	12	12	9	9	8	8	7	7	7	6	6	6	5	5	5	4	4	4	4	4	3	3	3	2	2	1	1	1
C01A7000169	5	1	4.7	21	15	14	13	12	11	10	10	10	9	9	8	8	8	7	7	6	6	6	6	5	5	5	5	4	4	4	3	3	2	1
C01A7000170	5	3	4.1	17	13	12	12	11	10	9	9	9	8	8	7	7	7	6	6	6	6	6	5	5	5	5	4	4	4	4	3	3	2	1
C01A7000171	5	1	3.6	13	11	10	10	9	8	8	7	7	6	6	6	6	6	5	5	5	5	5	4	4	4	4	3	3	3	3	3	2	2	2

Attachment G

Station Code	Year	Month	N	MAX	MIN	GMEAN	G.SD	ACTUAL VALUES						
70001	71	7	5	240.0	161.0	190.9	1.2	198.0	240.0	161.0	197.0	168.0		
70001	71	8	7	206.0	130.0	170.0	1.2	167.0	198.0	183.0	130.0	136.0	186.0	206.0
70001	71	9	6	168.0	91.0	134.9	1.3	120.0	168.0	160.0	122.0	91.0	168.0	
70001	71	10	6	181.0	72.0	131.9	1.4	166.0	142.0	72.0	128.0	181.0	134.0	
70001	71	11	6	304.0	66.0	168.7	1.7	179.0	304.0	66.0	200.0	249.0	129.0	
70001	71	12	5	158.0	29.0	87.2	1.9	96.0	158.0	106.0	108.0	29.0		
A70001	72	1	4	236.0	155.0	196.2	1.2	186.0	218.0	236.0	155.0			
A70001	72	2	6	302.0	103.0	180.9	1.4	162.0	160.0	103.0	302.0	184.0	236.0	
A70001	72	3	6	301.0	147.0	201.2	1.3	301.0	163.0	253.0	201.0	181.0	147.0	
A70001	72	4	5	175.0	124.0	140.9	1.2	141.0	124.0	125.0	175.0	145.0		
A70001	72	5	5	192.0	74.0	115.8	1.4	74.0	192.0	99.0	113.0	131.0		
A70001	72	6	5	170.0	119.0	135.9	1.2	170.0	119.0	122.0	136.0	138.0		
A70001	72	7	7	163.0	86.0	120.4	1.3	112.0	116.0	159.0	87.0	86.0	146.0	163.0
A70001	72	8	6	166.0	94.0	122.3	1.2	114.0	166.0	114.0	94.0	132.0	125.0	
A70001	72	9	6	150.0	62.0	103.1	1.3	94.0	62.0	150.0	102.0	108.0	125.0	
A70001	72	10	6	172.0	63.0	92.9	1.4	84.0	63.0	74.0	81.0	172.0	118.0	
A70001	72	11	6	223.0	70.0	115.8	1.6	223.0	135.0	86.0	84.0	70.0	158.0	
A70001	72	12	6	150.0	46.0	110.5	1.6	114.0	46.0	150.0	120.0	143.0	135.0	
A70060	71	7	5	231.0	137.0	196.3	1.2	193.0	218.0	137.0	219.0	231.0		
A70060	71	8	7	255.0	150.0	194.3	1.2	150.0	222.0	235.0	176.0	152.0	196.0	255.0
A70060	71	9	6	261.0	90.0	148.6	1.4	133.0	261.0	160.0	174.0	90.0	124.0	
A70060	71	10	6	164.0	92.0	112.8	1.3	164.0	92.0	92.0	126.0	99.0	119.0	
A70060	71	11	6	418.0	64.0	190.8	1.9	202.0	418.0	64.0	215.0	273.0	152.0	
A70060	71	12	6	120.0	28.0	71.5	1.7	120.0	104.0	74.0	99.0	28.0	52.0	
A70060	72	1	6	342.0	132.0	207.1	1.4	132.0	178.0	280.0	234.0	342.0	150.0	
A70060	72	2	6	318.0	84.0	163.3	1.6	134.0	130.0	84.0	318.0	175.0	233.0	
A70060	72	3	5	248.0	142.0	201.3	1.3	218.0	248.0	239.0	180.0	142.0		
A70060	72	4	6	180.0	115.0	151.1	1.2	155.0	177.0	115.0	127.0	165.0	180.0	
A70060	72	5	6	280.0	103.0	175.0	1.4	212.0	103.0	280.0	184.0	151.0	169.0	
A70060	72	6	5	272.0	132.0	198.6	1.3	272.0	132.0	196.0	236.0	186.0		
A70060	72	7	7	230.0	110.0	178.3	1.3	190.0	226.0	230.0	110.0	140.0	198.0	190.0
A70060	72	8	6	292.0	107.0	171.5	1.4	195.0	292.0	146.0	107.0	171.0	167.0	
PA70060	72	9	6	246.0	82.0	136.8	1.5	82.0	94.0	246.0	132.0	126.0	208.0	
PA70060	72	10	6	208.0	55.0	105.0	1.5	122.0	100.0	89.0	55.0	208.0	108.0	
PA70060	72	11	6	255.0	45.0	92.3	1.8	255.0	100.0	72.0	68.0	45.0	110.0	
PA70060	72	12	6	176.0	36.0	98.6	1.7	120.0	36.0	98.0	110.0	176.0	112.0	
PA70076	71	7	5	150.0	118.0	137.3	1.1	133.0	150.0	118.0	150.0	138.0		
PA70076	71	8	7	253.0	114.0	146.9	1.3	114.0	153.0	135.0	125.0	123.0	253.0	161.0
PA70076	71	9	6	182.0	97.0	128.0	1.3	102.0	151.0	117.0	138.0	97.0	182.0	
PA70076	71	10	6	187.0	108.0	145.4	1.2	187.0	160.0	108.0	152.0	122.0	158.0	
PA70076	71	11	6	276.0	66.0	166.6	1.7	142.0	226.0	66.0	203.0	276.0	180.0	
PA70076	71	12	6	199.0	42.0	110.6	1.7	136.0	199.0	137.0	108.0	42.0	109.0	
PA70076	72	1	6	303.0	128.0	184.4	1.3	303.0	177.0	195.0	166.0	177.0	128.0	
PA70076	72	2	6	229.0	97.0	150.5	1.4	229.0	159.0	133.0	221.0	112.0	97.0	
PA70076	72	3	6	179.0	109.0	139.6	1.2	179.0	113.0	159.0	141.0	150.0	109.0	
PA70076	72	4	6	165.0	137.0	150.7	1.1	146.0	154.0	165.0	143.0	161.0	137.0	
PA70076	72	5	6	166.0	97.0	137.9	1.2	129.0	97.0	154.0	138.0	166.0	156.0	
PA70076	72	6	6	162.0	137.0	145.6	1.1	141.0	144.0	153.0	137.0	162.0	138.0	
PA70076	72	7	7	143.0	71.0	108.6	1.3	106.0	111.0	103.0	71.0	102.0	143.0	142.0
PA70076	72	8	6	173.0	102.0	132.5	1.2	102.0	171.0	121.0	173.0	115.0	129.0	

Attachment H

Particulate Air Quality Data Base

Attachment G

I

Pollutant / Method / Station Code / Year / Month	N	MAX	MIN	GMEAN	G.SD	ACTUAL VALUES						
S04A7000173 1	6	11.3	4.5	7.8	1.4	8.3	6.1	10.8	8.0	11.3	4.5	
S04A7000173 2	6	10.0	3.1	5.1	1.5	4.4	4.3	4.9	6.3	3.1	10.0	
S04A7000173 3	6	12.2	5.1	7.0	1.4	7.8	6.2	5.9	12.2	6.4	5.1	
S04A7000173 4	6	50.3	2.4	13.0	3.7	2.4	21.2	37.5	18.7	2.7	50.3	
S04A7000173 5	6	72.2	6.1	18.2	2.7	6.1	9.7	42.0	72.2	23.2	8.7	
S04A7000173 6	6	52.0	5.2	12.7	2.2	5.2	16.7	11.4	7.7	10.8	52.0	
S04A7000173 7	7	42.6	16.7	29.4	1.4	42.2	35.7	42.6	23.4	16.7	24.5	31.1
S04A7000173 8	6	30.6	11.9	17.9	1.5	30.6	19.6	25.8	12.0	11.9	14.8	
S04A7000173 9	6	42.8	6.1	18.0	2.0	13.9	14.2	42.8	36.1	6.1	18.3	
S04A700017310	6	22.5	4.0	7.1	1.9	22.5	4.0	6.5	6.5	4.0	8.6	
S04A700017311	6	39.0	0.1	2.4	9.6	3.6	39.0	23.2	0.1	0.5	1.3	
S04A700017312	6	16.9	0.7	4.9	3.5	3.3	0.7	16.9	11.1	2.3	13.1	
S04A7000174 1	6	9.8	1.9	3.9	1.7	1.9	2.7	3.9	9.8	4.3	4.5	
S04A7000174 2	6	24.2	3.2	5.6	2.2	24.2	4.2	3.6	8.2	3.2	3.3	
S04A7000174 3	6	12.9	1.8	4.7	2.4	1.9	1.8	8.6	12.9	8.6	3.1	
S04A7000174 4	6	16.2	2.6	6.7	2.0	2.6	6.8	5.0	13.9	16.2	4.5	
S04A7000174 5	6	33.8	4.1	12.7	2.3	16.7	33.8	25.6	4.1	5.1	13.7	
S04A7000174 6	5	44.5	3.9	15.6	3.1	21.2	42.3	44.5	3.9	5.9		
S04A7000174 7	6	21.9	7.7	12.6	1.6	15.4	20.0	8.9	8.8	7.7	21.9	
S04A7000174 8	6	28.6	9.4	18.6	1.6	9.4	11.1	23.1	28.6	27.0	22.0	
S04A7000174 9	6	38.9	16.2	24.8	1.4	16.2	32.4	16.8	38.9	23.1	29.1	
S04A700017410	6	38.4	2.7	11.6	2.5	12.3	8.5	38.4	23.5	9.5	2.7	
S04A700017411	6	27.4	1.5	3.4	3.0	2.6	4.1	1.5	27.4	1.6	2.3	
S04A700017412	6	13.3	2.5	4.4	1.9	5.7	4.4	13.3	3.3	2.5	2.5	
S04A7000175 1	6	6.0	2.3	3.5	1.5	3.2	5.7	2.3	2.3	6.0	3.0	
S04A7000175 2	4	7.4	2.9	4.2	1.5	2.9	4.6	3.2	7.4			
S04A7000175 3	6	36.3	3.0	7.1	2.4	36.3	3.0	3.9	7.7	4.8	8.3	
S04A7000175 4	5	15.6	2.5	6.7	1.9	2.5	6.0	6.8	8.4	15.6		
S04A7000175 5	5	30.4	4.3	12.3	2.4	4.3	26.8	5.5	14.7	30.4		
S04A7000175 6	5	23.9	13.6	18.7	1.3	13.6	23.9	14.4	23.5	20.9		
S04A7000175 7	5	22.4	12.2	14.8	1.3	13.7	15.5	12.2	22.4	12.3		
S04A7000175 8	5	22.8	8.2	15.4	1.5	20.4	22.8	17.4	13.2	8.2		
S04A7000175 9	5	31.0	11.3	17.8	1.6	11.9	16.1	31.0	11.3	26.8		
S04A700017510	5	19.7	3.4	8.7	1.9	13.4	7.5	3.4	19.7	7.5		
S04A700017511	5	22.5	3.5	9.5	2.3	21.0	22.5	8.7	3.5	5.5		
S04A700017512	5	20.9	5.7	8.9	1.8	5.8	5.9	5.7	20.9	14.0		
S04A7000176 1	6	23.5	5.1	11.0	1.7	5.1	9.6	9.8	23.5	16.0	10.0	
S04A7000176 2	4	15.0	5.8	7.6	1.6	5.8	6.1	15.0	6.4			
S04A7000176 3	6	15.8	4.6	7.0	1.6	5.6	5.6	4.6	8.0	15.8	6.5	
S04A7000176 4	5	17.0	5.4	8.7	1.6	5.4	6.8	7.6	17.0	10.5		
S04A7000176 5	5	18.3	9.6	13.2	1.3	9.6	18.3	14.9	12.8	11.9		
S04A7000176 6	5	14.3	8.2	11.3	1.2	11.5	8.2	13.4	10.3	14.3		
S04A7000176 7	4	34.4	16.5	23.4	1.4	34.4	19.9	16.5	26.4			
S04A7000176 8	5	29.6	7.8	13.1	1.7	11.4	17.4	8.3	7.8	29.6		
S04A7000176 9	5	23.8	7.7	15.8	1.6	18.8	23.2	12.4	23.8	7.7		
S04A700017610	4	28.8	4.2	10.0	2.2	7.8	4.2	28.8	10.6			
S04A700017611	5	16.2	4.3	9.3	1.7	4.3	8.0	8.8	16.2	13.9		
S04A700017612	5	13.2	2.4	6.0	1.9	2.4	6.7	8.0	13.2	4.4		
S04A7007673 1	6	14.0	4.7	7.1	1.5	6.5	5.5	14.0	7.3	7.2	4.7	
S04A7007673 2	6	11.9	3.4	6.3	1.5	6.8	5.7	5.5	7.3	3.4	11.9	

Attachment I

Sulfate Air Quality Data Base

	MEAN	DOSE	F>10	E>10	F>15	E>15	F>20	E>20	F>25	E>25
OXIA7000173 1 1	1.3	967	0	0	0	0	0	0	0	0
OXIA7000173 2 1	1.4	940	0	0	0	0	0	0	0	0
OXIA7000173 3 1	1.7	1264	0	0	0	0	0	0	0	0
OXIA7000173 4 1	2.4	1728	3	1	0	0	0	0	0	0
OXIA7000173 5 1	3.4	2529	37	177	12	72	5	26	2	5
OXIA7000173 6 1	4.4	3168	90	609	32	339	19	217	14	132
OXIA7000173 7 1	4.0	2976	74	338	20	160	13	75	7	22
OXIA7000173 8 1	3.6	2678	52	94	5	3	0	0	0	0
OXIA7000173 9 1	3.6	2592	57	117	3	5	0	0	0	0
OXIA700017310 1	2.8	2083	29	77	6	18	1	1	0	0
OXIA700017311 1	1.9	1368	12	52	3	19	2	5	0	0
OXIA700017312 1	1.3	967	0	0	0	0	0	0	0	0
OXIA7000174 1 1	1.1	818	0	0	0	0	0	0	0	0
OXIA7000174 2 1	1.3	873	0	0	0	0	0	0	0	0
OXIA7000174 3 1	2.3	1711	22	96	7	18	0	0	0	0
OXIA7000174 4 1	3.0	2160	43	164	16	29	0	0	0	0
OXIA7000174 5 1	3.7	2752	29	111	7	26	1	2	0	0
OXIA7000174 6 1	4.5	3240	90	279	14	61	5	11	0	0
OXIA7000174 7 1	4.0	2976	74	265	18	58	3	3	0	0
OXIA7000174 8 1	4.5	3348	93	318	22	63	4	5	0	0
OXIA7000174 9 1	4.8	3456	126	540	50	128	7	9	0	0
OXIA700017410 1	3.0	2232	44	112	5	16	0	0	0	0
OXIA700017411 1	1.9	1368	3	0	0	0	0	0	0	0
OXIA700017412 1	1.7	1264	0	0	0	0	0	0	0	0
OXIA7000175 1 1	2.0	1488	11	32	1	0	0	0	0	0
OXIA7000175 2 1	1.8	1209	6	16	0	0	0	0	0	0
OXIA7000175 3 1	2.4	1785	14	36	1	1	0	0	0	0
OXIA7000175 4 1	2.5	1800	7	14	0	0	0	0	0	0
OXIA7000175 5 1	3.5	2604	37	64	1	0	0	0	0	0
OXIA7000175 6 1	3.7	2664	36	79	2	4	0	0	0	0
OXIA7000175 7 1	4.6	3422	111	310	22	48	2	5	0	0
OXIA7000175 8 1	4.2	3124	93	265	22	45	3	0	0	0
OXIA7000175 9 1	4.6	3312	90	389	36	99	7	9	0	0
OXIA700017510 1	2.8	2083	37	148	14	26	1	1	0	0
OXIA700017511 1	2.1	1512	7	14	0	0	0	0	0	0
OXIA700017512 1	1.3	967	0	0	0	0	0	0	0	0
OXIA7000176 1 1	2.4	1785	11	19	0	0	0	0	0	0
OXIA7000176 2 1	2.0	1344	6	9	0	0	0	0	0	0
OXIA7000176 3 1	2.7	2008	13	50	3	3	0	0	0	0
OXIA7000176 4 1	2.7	1944	14	39	3	2	0	0	0	0
OXIA7000176 5 1	3.6	2678	37	152	12	42	3	6	0	0
OXIA7000176 6 1	4.0	2880	72	293	21	95	7	27	2	8
OXIA7000176 7 1	4.0	2976	67	193	11	31	1	5	0	0
OXIA7000176 8 1	3.3	2455	67	261	22	71	3	19	1	6
OXIA7000176 9 1	3.2	2304	43	151	9	66	4	36	3	17
OXIA700017610 1	3.1	2306	74	280	22	61	3	10	0	0
OXIA700017611 1	3.3	2376	43	127	7	16	0	0	0	0
OXIA700017612 1	1.9	1413	0	0	0	0	0	0	0	0

Attachment J. Oxidant Exposure Parameter Data Base.

	MEAN	DOSE	F>5	E>5	F>15	E>15	F>20	E>20	F>25	E>25
NO2A7000173 1 1	8.1	6026	669	2496	37	215	14	104	9	46
NO2A7000173 2 1	6.0	4032	403	1039	6	11	0	0	0	0
NO2A7000173 3 1	4.4	3273	297	540	0	0	0	0	0	0
NO2A7000173 4 1	5.5	3960	360	1017	12	40	2	3	0	0
NO2A7000173 5 1	6.9	5133	520	1740	13	83	6	34	3	12
NO2A7000173 6 1	6.8	4896	504	1699	25	126	10	41	3	8
NO2A7000173 7 1	6.0	4464	446	1206	14	53	4	7	0	0
NO2A7000173 8 1	6.1	4538	446	1241	14	77	6	30	2	8
NO2A7000173 9 2	6.1	4392	360	1377	28	57	3	2	0	0
NO2A700017310 1	9.0	6696	483	3414	111	763	48	393	32	195
NO2A700017311 1	6.7	4824	432	1683	28	89	7	6	0	0
NO2A700017312 1	8.2	6100	595	2588	44	97	5	9	0	0
NO2A7000174 1 1	7.9	5877	520	2420	67	257	18	66	5	9
NO2A7000174 2 1	7.2	4838	504	1823	11	43	3	5	0	0
NO2A7000174 3 1	4.5	3348	297	550	4	13	0	0	0	0
NO2A7000174 4 1	6.4	4608	432	1458	12	58	5	16	0	0
NO2A7000174 5 1	6.4	4761	520	1346	6	45	3	19	2	5
NO2A7000174 6 1	4.8	3456	288	954	18	100	7	41	3	18
NO2A7000174 7 1	6.6	4910	520	1600	22	108	7	38	3	17
NO2A7000174 8 1	6.0	4464	446	1272	22	83	7	13	0	0
NO2A7000174 9 1	7.8	5616	576	2249	32	258	18	137	10	68
NO2A700017410 1	9.1	6770	595	3224	83	550	29	263	12	172
NO2A700017411 1	11.8	8496	684	4842	162	783	50	316	18	180
NO2A700017412 1	11.3	8407	632	4922	167	1221	93	598	48	266
NO2A7000175 1 1	11.1	8258	669	4616	148	1203	93	617	44	282
NO2A7000175 2 1	7.8	5241	470	2118	50	289	18	123	10	52
NO2A7000175 3 1	6.1	4538	446	1270	11	35	3	8	0	0
NO2A7000175 4 1	5.0	3600	360	673	6	26	1	4	0	0
NO2A7000175 5 1	4.7	3496	334	600	5	7	0	0	0	0
NO2A7000175 6 1	4.5	3240	252	614	3	18	2	4	0	0
NO2A7000175 7 1	4.6	3422	260	784	14	40	2	5	0	0
NO2A7000175 8 1	4.5	3348	223	732	7	42	3	22	2	10
NO2A7000175 9 1	5.4	3888	288	1184	25	192	12	110	9	56
NO2A700017510 1	6.1	4538	446	1337	14	117	9	56	3	22
NO2A700017511 1	8.2	5904	504	2554	72	448	28	214	14	124
NO2A700017512 1	12.2	9076	595	5573	204	1583	111	830	52	454
NO2A7000176 1 1	8.6	6398	669	2737	37	87	5	6	0	0
NO2A7000176 2 1	6.2	4166	336	1249	20	97	6	37	4	9
NO2A7000176 3 1	6.7	4984	520	1620	22	63	5	4	0	0
NO2A7000176 4 1	5.7	4104	432	891	7	28	3	7	0	0
NO2A7000176 5 1	5.4	4017	372	835	9	58	3	26	2	10
NO2A7000176 6 1	6.6	4752	504	1481	28	145	11	58	3	22
NO2A7000176 7 1	5.6	4166	446	965	13	41	2	3	0	0
NO2A7000176 8 1	6.4	4761	372	1591	37	293	18	159	9	89
NO2A7000176 9 1	6.6	4752	504	1355	19	51	1	1	0	0
NO2A700017610 1	9.2	6844	595	3324	93	460	29	218	14	114
NO2A700017611 1	10.5	7560	612	4096	144	649	50	262	19	105
NO2A700017612 1	10.1	7514	669	3757	111	325	22	63	1	0

Attachment K. NO_2 Exposure Parameter Data Base

		MEAN	DOSE	F>5	E>5	F>10	E>10	F>15	E>15	F>20	E>20
C01A7000173 1	1	5.6	4166	372	1096	74	259	14	54	5	13
C01A7000173 2	1	4.9	3292	302	672	40	83	3	5	0	0
C01A7000173 3	1	3.7	2752	186	249	7	5	0	0	0	0
C01A7000173 4	1	4.1	2952	216	338	14	18	0	0	0	0
C01A7000173 5	1	4.0	2976	223	276	7	14	0	0	0	0
C01A7000173 6	1	4.6	3312	288	505	28	66	5	9	0	0
C01A7000173 7	1	4.0	2976	204	226	7	8	0	0	0	0
C01A7000173 8	1	4.2	3124	223	273	11	16	0	0	0	0
C01A7000173 9	1	4.6	3312	252	539	36	96	5	3	0	0
C01A700017310	1	6.5	4836	372	1748	148	539	44	146	11	37
C01A700017311	1	5.6	4032	360	1101	72	210	14	19	0	0
C01A700017312	1	7.1	5282	520	1877	148	501	37	92	3	5
C01A7000174 1	1	6.4	4761	446	1661	130	378	29	90	3	3
C01A7000174 2	1	4.9	3292	268	793	53	157	13	26	0	0
C01A7000174 3	1	3.3	2455	130	303	22	44	3	6	0	0
C01A7000174 4	1	2.9	2088	108	108	3	2	0	0	0	0
C01A7000174 5	1	2.4	1785	59	63	1	0	0	0	0	0
C01A7000174 6	1	2.9	2088	108	152	7	18	1	0	0	0
C01A7000174 7	1	3.8	2827	204	261	7	6	0	0	0	0
C01A7000174 8	1	3.4	2529	167	232	7	3	0	0	0	0
C01A7000174 9	1	5.6	4032	360	968	64	171	14	30	1	3
C01A700017410	1	5.5	4092	372	1039	67	253	22	85	7	22
C01A700017411	1	7.5	5400	504	2193	180	690	57	179	10	44
C01A700017412	1	8.5	6324	520	2919	223	1115	93	361	29	77
C01A7000175 1	1	8.4	6249	520	2863	223	1096	93	380	29	123
C01A7000175 2	1	6.0	4032	336	1258	100	342	23	80	6	18
C01A7000175 3	1	4.5	3348	260	653	44	97	7	19	0	0
C01A7000175 4	1	3.2	2304	144	270	12	39	4	4	0	0
C01A7000175 5	1	3.2	2380	148	195	7	14	0	0	0	0
C01A7000175 6	1	3.3	2376	162	243	10	34	3	6	0	0
C01A7000175 7	1	2.6	1934	74	82	3	8	0	0	0	0
C01A7000175 8	1	3.1	2306	111	197	7	16	0	0	0	0
C01A7000175 9	1	4.1	2952	252	462	28	102	9	24	1	0
C01A700017510	1	4.4	3273	260	644	37	121	9	26	1	2
C01A700017511	1	5.9	4248	360	1428	108	429	36	121	10	35
C01A700017512	1	7.0	5208	446	2043	186	564	52	99	3	10
C01A7000176 1	1	6.7	4984	446	1743	148	423	29	73	5	7
C01A7000176 2	1	4.0	2688	201	509	26	78	5	11	0	0
C01A7000176 3	1	3.8	2827	223	350	14	19	0	0	0	0
C01A7000176 4	1	3.3	2376	144	173	5	10	0	0	0	0
C01A7000176 5	1	2.5	1860	74	93	0	0	0	0	0	0
C01A7000176 6	1	3.2	2304	126	180	10	30	1	3	0	0
C01A7000176 7	1	3.0	2232	74	74	1	0	0	0	0	0
C01A7000176 8	1	3.3	2455	148	349	22	76	4	16	1	0
C01A7000176 9	1	3.2	2304	144	221	7	14	0	0	0	0
C01A700017610	1	4.3	3199	260	564	29	56	2	7	0	0
C01A700017611	1	6.2	4464	360	1477	126	352	28	52	4	7
C01A700017612	1	6.7	4984	446	1743	148	423	22	62	1	0

Attachment L. CO Exposure Parameter Data Base

	MEAN	DOSE	F>2	E>2	F>4	E>4	F>8	E>8	F>14	E>14
S02A7000173 1 1	1.5	1116	260	95	7	0	0	0	0	0
S02A7000173 2 1	1.6	1075	302	88	6	2	0	0	0	0
S02A7000173 3 1	1.3	967	167	31	3	0	0	0	0	0
S02A7000173 4 1	1.4	1008	198	81	14	2	0	0	0	0
S02A7000173 5 1	2.3	1711	446	473	111	110	5	3	0	0
S02A7000173 6 1	2.0	1440	360	314	64	56	1	0	0	0
S02A7000173 7 1	2.4	1785	520	479	93	60	1	0	0	0
S02A7000173 8 1	2.2	1636	372	433	93	89	1	0	0	0
S02A7000173 9 1	2.1	1512	360	373	90	40	0	0	0	0
S02A700017310 1	2.0	1488	372	338	74	59	1	0	0	0
S02A700017311 1	2.1	1512	360	349	64	109	7	16	0	0
S02A700017312 1	1.8	1339	372	191	29	18	0	0	0	0
S02A7000174 1 1	1.6	1190	297	125	14	4	0	0	0	0
S02A7000174 2 1	1.6	1075	336	93	13	0	0	0	0	0
S02A7000174 3 1	1.4	1041	223	48	5	0	0	0	0	0
S02A7000174 4 1	1.4	1008	198	42	1	0	0	0	0	0
S02A7000174 5 1	1.4	1041	223	75	7	0	0	0	0	0
S02A7000174 6 1	2.1	1512	432	330	72	42	0	0	0	0
S02A7000174 7 1	1.9	1413	446	185	29	12	0	0	0	0
S02A7000174 8 1	1.7	1264	334	149	37	6	0	0	0	0
S02A7000174 9 1	2.2	1584	504	348	72	33	0	0	0	0
S02A700017410 1	1.6	1190	334	133	37	9	0	0	0	0
S02A700017411 1	2.4	1728	576	429	108	42	0	0	0	0
S02A700017412 1	2.4	1785	520	464	130	82	1	0	0	0
S02A7000175 1 1	2.4	1785	520	503	130	121	3	9	0	0
S02A7000175 2 1	2.1	1411	403	288	67	52	5	3	0	0
S02A7000175 3 1	1.6	1190	297	103	22	14	0	0	0	0
S02A7000175 4 1	1.3	936	216	13	1	0	0	0	0	0
S02A7000175 5 1	1.8	1339	372	159	29	4	0	0	0	0
S02A7000175 6 1	1.4	1008	252	32	3	2	0	0	0	0
S02A7000175 7 1	1.7	1264	297	187	44	18	0	0	0	0
S02A7000175 8 1	1.9	1413	446	183	29	10	0	0	0	0
S02A7000175 9 1	2.2	1584	504	363	72	48	0	0	0	0
S02A700017510 1	1.7	1264	334	146	14	6	0	0	0	0
S02A700017511 1	2.3	1656	504	361	72	46	0	0	0	0
S02A700017512 1	3.5	2604	669	1141	334	229	0	0	0	0
S02A7000176 1 1	2.7	2008	595	630	167	100	1	3	0	0
S02A7000176 2 1	1.4	940	268	147	20	14	0	0	0	0
S02A7000176 3 1	1.9	1413	297	484	93	224	22	2	0	0
S02A7000176 4 1	1.0	720	180	54	5	1	0	0	0	0
S02A7000176 5 1	1.4	1041	297	181	44	12	0	0	0	0
S02A7000176 6 1	1.6	1152	360	170	28	20	1	2	0	0
S02A7000176 7 1	2.3	1711	520	462	111	62	0	0	0	0
S02A7000176 8 1	2.0	1488	446	348	74	23	0	0	0	0
S02A7000176 9 2	2.0	1440	432	290	50	2	0	0	0	0
S02A700017610 1	2.0	1488	446	274	52	23	0	0	0	0
S02A700017611 1	2.3	1656	504	433	90	64	1	0	0	0
S02A700017612 1	2.6	1934	669	522	93	11	0	0	0	0

Attachment M. SO_2 Exposure Parameter Data Base

	MEAN	DOSE	F>60	E>60	F>100	E>100	F>150	E>150	F>200	E>200
TSPA7000173 1	1128.4	95529	744	1591	428	1000	318	815	111	751
TSPA7000173 2	1 85.2	57254	463	1041	332	771	0	0	0	0
TSPA7000173 3	1 88.5	65844	600	1164	331	825	98	756	0	0
TSPA7000173 4	1115.4	83088	720	1384	469	912	300	730	0	0
TSPA7000173 5	1139.5	103788	744	1729	635	1043	173	828	126	781
TSPA7000173 6	1108.9	78408	648	1311	415	886	225	740	0	0
TSPA7000173 7	1114.1	84890	744	1414	584	871	0	0	0	0
TSPA7000173 8	1128.3	95455	744	1590	396	1022	175	873	142	821
TSPA7000173 9	1114.1	82152	631	1390	486	923	180	751	0	0
TSPA700017310	1146.4	108921	744	1815	744	1089	310	816	117	750
TSPA700017311	1102.1	73512	571	1281	268	973	200	852	155	794
TSPA700017312	1118.2	87940	744	1466	487	930	214	770	0	0
TSPA7000174 1	1 76.5	56916	566	1092	216	812	90	752	0	0
TSPA7000174 2	1126.2	84806	672	1413	600	851	0	0	0	0
TSPA7000174 3	1 59.3	44119	369	933	149	783	0	0	0	0
TSPA7000174 4	1 87.4	62928	599	1097	333	752	0	0	0	0
TSPA7000174 5	1 94.0	69936	744	1165	353	780	0	0	0	0
TSPA7000174 6	1110.3	79416	720	1324	462	852	0	0	0	0
TSPA7000174 7	3112.4	83662	744	1393	541	866	0	0	0	0
TSPA7000174 8	1 93.2	69340	744	1155	418	791	0	0	0	0
TSPA7000174 9	1146.7	105624	720	1760	720	1056	420	756	0	0
TSPA700017410	1 99.0	73656	652	1236	346	865	151	755	0	0
TSPA700017411	1122.6	88272	720	1471	509	920	227	752	0	0
TSPA700017412	1100.3	74623	591	1349	441	937	154	806	106	767
TSPA7000175 1	1105.9	78789	744	1313	391	855	120	767	0	0
TSPA7000175 2	2 75.7	50870	484	882	164	695	0	0	0	0
TSPA7000175 3	1 87.9	65397	520	1166	288	871	203	762	0	0
TSPA7000175 4	1 74.7	53784	533	1028	243	773	0	0	0	0
TSPA7000175 5	1117.2	87196	744	1453	558	927	190	760	0	0
TSPA7000175 6	1110.4	79488	720	1324	460	821	0	0	0	0
TSPA7000175 7	1110.8	82435	744	1373	498	862	0	0	0	0
TSPA7000175 8	1 89.8	66811	744	1114	252	747	0	0	0	0
TSPA7000175 9	1105.8	76176	720	1269	402	836	110	724	0	0
TSPA700017510	1102.8	76483	744	1274	223	809	79	745	0	0
TSPA700017511	1144.9	104328	720	1738	720	1043	216	787	123	736
TSPA700017512	1177.5	132060	744	2201	744	1320	584	900	184	777
TSPA7000176 1	1138.5	103044	662	1741	590	1132	409	847	206	749
TSPA7000176 2	2 86.1	57859	472	1050	249	791	147	701	0	0
TSPA7000176 3	1 86.8	64579	572	1188	376	855	95	749	0	0
TSPA7000176 4	1 82.4	59328	462	1093	367	814	102	726	0	0
TSPA7000176 5	1 99.3	73879	744	1230	352	845	0	0	0	0
TSPA7000176 6	1107.7	77544	720	1292	360	814	0	0	0	0
TSPA7000176 7	2114.0	84816	744	1413	426	907	195	758	0	0
TSPA7000176 8	1 86.8	64579	537	1131	372	830	0	0	0	0
TSPA7000176 9	1120.1	86472	720	1441	453	971	330	775	0	0
TSPA700017610	1 97.9	72837	744	1214	346	845	104	748	0	0
TSPA700017611	1 99.5	71640	720	1193	431	795	0	0	0	0
TSPA700017612	1120.1	89354	744	1488	614	900	122	754	0	0

Attachment N. TSP Exposure Parameter Data Base

	MEAN	DOSE	F>10	E>10	F>15	E>15	F>20	E>20	F>25	E>25
SO4A7000173 1 1	7.8	5803	222	765	0	0	0	0	0	0
SO4A7000173 2 1	5.1	3427	56	672	0	0	0	0	0	0
SO4A7000173 3 1	7.0	5208	117	762	0	0	0	0	0	0
SO4A7000173 4 1	13.0	9360	458	1426	433	1109	355	939	265	857
SO4A7000173 5 1	18.2	13540	429	1508	372	1212	331	1058	294	962
SO4A7000173 6 1	12.7	9144	447	1063	213	896	160	833	137	796
SO4A7000173 7 1	29.4	21873	744	2189	744	1459	634	1116	469	938
SO4A7000173 8 1	17.9	13317	744	1330	428	957	300	832	200	771
SO4A7000173 9 1	18.0	12960	588	1384	394	1022	284	896	244	825
SO4A700017310 1	7.1	5282	166	842	114	780	77	752	0	0
SO4A700017311 1	2.4	1728	234	1000	208	883	189	815	162	771
SO4A700017312 1	4.9	3645	320	857	120	754	0	0	0	0
SO4A7000174 1 1	3.9	2901	0	0	0	0	0	0	0	0
SO4A7000174 2 1	5.6	3763	147	768	105	711	75	684	0	0
SO4A7000174 3 1	4.7	3496	139	770	0	0	0	0	0	0
SO4A7000174 4 1	6.7	4824	235	812	120	727	0	0	0	0
SO4A7000174 5 1	12.7	9448	473	1202	377	944	257	838	192	784
SO4A7000174 6 1	15.6	11232	444	1459	398	1150	366	987	325	886
SO4A7000174 7 1	12.6	9374	407	1014	316	833	186	755	0	0
SO4A7000174 8 1	18.6	13838	635	1391	503	1032	451	858	247	768
SO4A7000174 9 1	24.8	17856	720	1782	720	1188	474	944	378	826
SO4A700017410 1	11.6	8630	409	1107	272	927	216	843	170	795
SO4A700017411 1	3.4	2448	123	818	98	769	79	742	65	725
SO4A700017412 1	4.4	3273	103	768	0	0	0	0	0	0
SO4A7000175 1 1	3.5	2604	0	0	0	0	0	0	0	0
SO4A7000175 2 2	4.2	2822	0	0	0	0	0	0	0	0
SO4A7000175 3 1	7.1	5282	170	909	136	837	112	797	93	773
SO4A7000175 4 1	6.7	4824	175	777	81	723	0	0	0	0
SO4A7000175 5 1	12.3	9151	430	1194	367	960	295	845	240	779
SO4A7000175 6 1	18.7	13464	720	1347	488	920	377	774	0	0
SO4A7000175 7 1	14.8	11011	744	1102	262	809	120	755	0	0
SO4A7000175 8 1	15.4	11457	607	1181	451	880	241	765	0	0
SO4A7000175 9 1	17.8	12816	720	1283	393	943	298	823	235	757
SO4A700017510 1	8.7	6472	298	890	179	779	0	0	0	0
SO4A700017511 1	9.5	6840	337	970	271	818	224	741	0	0
SO4A700017512 1	8.9	6621	281	903	197	790	90	747	0	0
SO4A7000176 1 1	11.0	8184	310	927	203	806	114	758	0	0
SO4A7000176 2 2	7.6	5107	164	724	84	672	0	0	0	0
SO4A7000176 3 1	7.0	5208	145	793	71	747	0	0	0	0
SO4A7000176 4 1	8.7	6264	237	805	109	731	0	0	0	0
SO4A7000176 5 1	13.2	9820	641	985	218	773	0	0	0	0
SO4A7000176 6 1	11.3	8136	522	846	0	0	0	0	0	0
SO4A7000176 7 2	23.4	17409	744	1738	744	1159	461	912	314	812
SO4A7000176 8 1	13.1	9746	433	1051	275	869	184	796	121	760
SO4A7000176 9 1	15.8	11376	569	1186	438	892	317	764	0	0
SO4A700017610 2	10.0	7440	314	977	214	851	160	791	119	759
SO4A700017611 1	9.3	6696	319	839	144	728	0	0	0	0
SO4A700017612 1	6.0	4464	156	776	0	0	0	0	0	0

Attachment O. Sulfate Exposure Parameter Data Base

Attachment P

Cubic Spline Interpolation Scheme With A Parabolic Leapfrog Correction

Reference

1. Sommers W.T. (1976), "Data Interpolation by Cubic Splines", USDA Forest Service Research Note PSW - 313

2. Guardado, J.L. and W.T. Sommers (1977), "Interpolation of Unevenly Spaced Data Using a Parabolic Leapfrog Correction Method and Cubic Splines", USDA Forest Service Research Note PSW - 324.

PACIFIC SOUTHWEST
Forest and Range
Experiment Station

FOREST SERVICE
U.S.DEPARTMENT OF AGRICULTURE
P.O. BOX 245, BERKELEY, CALIFORNIA 94701

DATA INTERPOLATION BY CUBIC SPLINES

William T. Sommers

USDA Forest Service
Research Note PSW-313
1976

Sommers, William T.
1976. **Data interpolation by cubic splines.** USDA Forest Serv. Res. Note PSW-313, 5 p., illus. Pacific Southwest Forest and Range Exp. Stn., Berkeley, Calif.

The cubic spline is a cubic polynomial that can be used with data sets of either fixed or variable mesh length. It is useful for interpolating data, smoothing data, estimating volume, and achieving greater precision in numerical analysis. The technique has a variety of applications in forestry and related disciplines. A FORTRAN IV computer program, available upon request, will yield a bi-cubic spline interpolation for either a variable mesh length or an equal interval grid.

Oxford: 0—015.5
Retrieval Terms: data smoothing; forestry; fire management; cubic splines.

Cubic splines are a mathematical formulation of a mechanical technique used by draftsmen for many years, in which they fit a flexible strip of material to known points in order to form a smooth curve. For mathematical purposes, the known points are data values on a fixed or variable mesh and the flexible strip is approximated by a piecewise continuous function. This function is defined as a cubic polynomial in each interval between the known points, with continuous first and second derivations.

The cubic spline is useful as a data interpolation technique, as a finite element method in numerical analysis, as a data smoother and as a volume estimator. Meteorologists have used cubic splines for data smoothing[1], for telescoping a grid[2], and for evaluating derivatives in numerical analysis computations.[3] Cubic splines are of potential use in forestry and fire applications that deal with numeric data bases. Examples of such use include smoothing or interpolating digital topography, working with brush cover heights and classes, estimating volume contained between tree tops and ground level from sparse data, interpolating growth rates between known points, and in other applications dealing with a spatial or time series data base.

SINGLE CUBIC SPLINE

The basic mathematical theory for both the single cubic spline (one-dimensional) and bi-cubic spline (two-dimensional) is explained by Ahlberg and others.[4] A concise reformulation of the basic theory tailored to application follows.

If $S_d(f;x)$ is the spline of interpolation of the function $f(x)$ on the mesh d, then

$$\int_a^b \left| f''(x) \right|^2 dx =$$

$$\int_a^b \left| s_d'' (f;x) \right|^2 dx + \int_a^b \left| f''(x) - s_d'' (f;x) \right|^2 dx \qquad (1)$$

where primes represent differentiation. The integral on the left hand side of (1) is minimized by the spline fit and end conditions $s_d''(f;a) = s_d''(f;b) = 0$. In words, the cubic spline represents the curve with minimum curvature that will fit the given data, be continuous through the second derivative and assure the "smoothest" fit to the data.

The derivations which follow may be found in expanded form in the publication by Ahlberg and others[4/]. Consider the mesh d of length b-a sub-divided such that

$d: a = x_0 < x_i < \ldots < x_n = b$

where the spacing between the x_i need not be uniform. Data values at the x_i points are given by

$Z: z_0, z_i, \ldots z_n$

and may represent digital elevation values or some other form of data on a grid. We seek a spline function $S_d(x)$ which is continuous through its second derivative ($S_d''(x)$), coincides with a cubic in each sub-interval $x_{i-1} \le x \le x_i$ ($i = 1, 2, \ldots n$) and satisfies $S_d(x_i) = z_i$ at the given points x_i. $S_d(x)$ is periodic or nonperiodic depending on whether the condition $S_d(a+) = S_d(b-)$ is satisfied or not.

We designate M_i the moment S_d'' and from the prescribed linearity of the second derivative we write

$$S''_d(x) = M_{i-1}\left(\frac{x_i-x}{h_i}\right) + M_i\left(\frac{x-x_{i-1}}{h_i}\right) \qquad (2)$$

where $h_i = x_i - x_{i-1}$ is the spacing between the known data points. Equation 2 simply restates the minimum curvature property and uses the prescribed linear change of the second derivative between known points x_i.

Integrating (2) twice and evaluating the constants yields

$$S_d'(x) = -M_{i-1}\left(\frac{(x_i-x)^2}{2h_i}\right) + M_i\left(\frac{(x-x_{i-1})^2}{2h_i}\right) +$$

$$\left(\frac{z_i-z_{i-1}}{h_i}\right)\left(\frac{M_i-M_{i-1}}{6}\right)h_i \qquad (3)$$

$$S_d(x) = M_{i-1}\left(\frac{(x_i-x)^3}{6h_i}\right) + M_i\left(\frac{(x-x_{i-1})^3}{6h_i}\right) +$$

$$\left(z_{i-1} - \left(\frac{M_{i-1}\, h_i^2}{6}\right)\right)\left(\frac{x_i-x}{h_i}\right) +$$

$$\left(z_i - \left(\frac{M_i h_i^2}{6}\right)\right)\left(\frac{x-x_{i-1}}{h_i}\right) \qquad (4)$$

Equations 3 give the slope of the spline and equations 4 the spline itself. We will continue the development here in terms of the moments, M_i, but one could just as easily develop the technique in terms of the slopes. One sided limits of (3) yield

$$S_d'(x_i-) = \left(\frac{h_i}{6} M_{i-1}\right) + \left(\frac{h_i}{3} M_i\right) + \left(\frac{z_i-z_{i-1}}{h_i}\right) \qquad (5a)$$

$$S_d'(x_i+) = -\left(\frac{h_{i+1}}{3} M_i\right) - \left(\frac{h_{i+1}}{6} M_{i+1}\right) + \left(\frac{z_{i+1}-z_i}{h_{i+1}}\right) \qquad (5b)$$

These and the slope continuity condition yield

$$\left(\frac{h_i}{6} M_{i-1}\right) + \left(\frac{h_i+h_{i+1}}{3} M_i\right) + \left(\frac{h_{i+1}}{6} M_{i+1}\right) =$$

$$\left(\frac{z_{i+1}-z_i}{h_{i+1}}\right) - \left(\frac{z_i-z_{i-1}}{h_i}\right) \qquad (6)$$

which, for the periodic spline, give N-1 simultaneous equations for the moments $M_i, M_2, \ldots M_n$. For the nonperiodic spline, additional, or end, conditions must be supplied. These conditions are a matter of experiment. We have found that specifying the slope at the end points (a,b) by linearly extrapolating the z values from (z_1, z_2) and (z_{n-1}, z_n) works well. Generalized end conditions are

$$2M_0 + \lambda_0 M_1 = C_0 \qquad (7a)$$

$$\mu_n M_{n-1} + 2M_n = C_n \qquad (7b)$$

in which

$$\lambda_i = \frac{h_{i+1}}{h_i+h_{i+1}} \qquad (8a)$$

$$\mu_i = 1-\lambda_i \quad (i = 1, 2, \ldots, n-1) \qquad (8b)$$

Equations 6 now become

$$\mu_i M_{i-1} + 2M_i + \lambda_i M_{i+1} = 6$$

$$\frac{\left[(z_{i+1}-z_1)/h_{i+1}\right] - \left[(z_i-z_{i-1})/h_i\right]}{h_i+h_{i+1}} \qquad (9)$$

Writing equations 7 and 8 in matrix form as

$$\begin{bmatrix} 2 & \lambda_o & 0 & \cdots & 0 & 0 & 0 \\ \mu_i & 2 & \lambda_i & \cdots & 0 & 0 & 0 \\ 0 & \mu_2 & 2 & \cdots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 2 & \lambda_{n-2} & 0 \\ 0 & 0 & 0 & \cdots & \mu_{n-1} & 2 & \lambda_{n-1} \\ 0 & 0 & 0 & \cdots & 0 & \mu_n & 2 \end{bmatrix} \begin{bmatrix} M_o \\ M_1 \\ M_2 \\ \cdot \\ \cdot \\ \cdot \\ M_{n-2} \\ M_{n-1} \\ M_n \end{bmatrix} = \begin{bmatrix} C_o \\ C_1 \\ C_2 \\ \cdot \\ \cdot \\ \cdot \\ C_{n-2} \\ C_{n-1} \\ C_n \end{bmatrix} \tag{10a}$$

or $AM = C$ (10b)

shows them to be a tri-diagonal system of equations which can be solved by numerical techniques. Repeated calculations on the same mesh (i.e., only the z_i change) can yield considerable time savings due to the non-variance of h_i, λ_i, and μ_i.

Use of equal interval meshes can save even more computational time. In many meteorological applications, large iterative computations are carried out on equal interval, multi-level grids. As a result much time can be saved. For an equal interval mesh, the inverse of the coefficient matrix A of equations 10 takes on simplified form. Equation 10b may be rewritten as $M = A^{-1}C$, and the moments then may be calculated by inverting the coefficient matrix A. The nxn determinant of A in terms of constant λ is

$$D_n(\lambda) \begin{vmatrix} 2 & \lambda & 0\ldots & & & \\ 1-\lambda & 2 & \lambda & & & \\ 0 & \lambda & 2 & & & \\ & & \cdots & 2 & 0 & 0 \\ & & & 1-\lambda & 2 & \lambda \\ & & & 0 & 1-\lambda & 2 \end{vmatrix} \tag{11}$$

since $\lambda = \dfrac{h_i}{h_i + h_{i+1}} = 1/2$ for equal intervals, the determinant $D_n = D_n(1/2)$ yields

$$\left.\begin{aligned} D_n &= \frac{\left(\frac{1+3^{1/2}}{2}\right)^{n+1} - \left(\frac{1-3^{1/2}}{2}\right)^{n+1}}{3^{1/2}} \\ D_o &= 1 \\ D_{-1} &= 0 \end{aligned}\right\} \tag{12}$$

The coefficient determinant $|A|$ satisfies

$$|A| = 4D_{n-1} - 1/2D_{n-2} \tag{13}$$

And the elements of the inverse of the coefficient matr
are

$$A^{-1}_{i,j} = \frac{(-1)^{i+j}\left(2D_{j-1} - 1/4D_{i-2}\right)\left(2D_{n-j-1} - 1/4D_{n-j-2}\right)}{2^{i-j}\ |A|}$$

$(0 < j \leq i \leq N)$ (14a

$$A^{-1}_{i,j} = \frac{(-1)^{i+j}\left(2D_{j-1} - 1/4D_{j-2}\right)\left(2D_{n-i-1} - 1/4D_{n-i-2}\right)}{2^{i-j}\ |A|}$$

$(0 \leq j \leq i < N)$ (14b

$$A^{-1}_{o,j} = \frac{(-1)^{j}\ 1/2\left(2D_{n-j-1} - 1/4D_{n-j-2}\right)}{2^{j-1}\ |A|}$$

$(0 < j \leq N)$ (14c

$$A^{-1}_{N,j} = \frac{(-1)^{N+j}\ 1/2\left(2D_{j-1} - 1/4D_{j-2}\right)}{2^{N-j-1}\ |A|}$$

$(0 < j < N)$ (14d

Since $M_i = \sum_{j=o}^{N} A^{-1}_{i,j}\ C_j$

we get

$$M_i = A^{-1}_{i,o}\left[C_o + \frac{6}{h^2}\left(Z_o + \lambda_o Z_1\right)\right] - \frac{18}{h^2}\sum_{j=o}^{N} A^{-1}_{i,j}\ Z_j + \frac{6}{h^2} Z_i + A^{-1}_{i,N}\left[C_N + \frac{6}{h^2}\left(Z_N - \mu_N Z_{N-1}\right)\right] \tag{16}$$

Additional shortcuts can be used when the spline is use operationally on an equal interval mesh when the Z fiel does not vary drastically from run to run. Since the $A^{-1}_{i,j}$ decay rapidly with distance from the principal diagonal (i.e., abs(i-j) becomes large), experimentatio should yield suitable truncation methods.

Bi-cubic Spline

We have dealt so far with the single cubic spline. For a bi-cubic spline, extension to a grid of k rows by columns is easily accomplished by considering the calcu lated $S_d(x)$ values to be new z values and rotating the single cubic spline technique through 90°. For example, consider that you wish to interpolate to a point κ, γ between rows k and k+1 and columns l and l+1. Calculate $S_d(\gamma)$ for all rows and calculate $S_d(\kappa, \gamma)$ for the interpolated values $S_d(\gamma)$ in the "γ column".

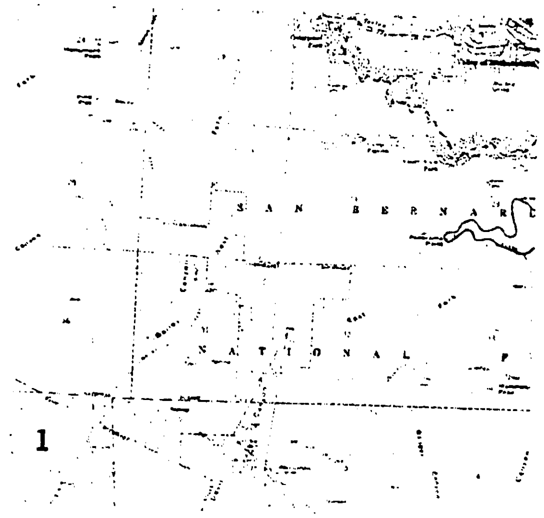


Figure 1—Section of USGS San Bernardino north quadrangle 7.5-minute topographic map shows the Devil Canyon area of the San Bernardino National Forest, in southern California.

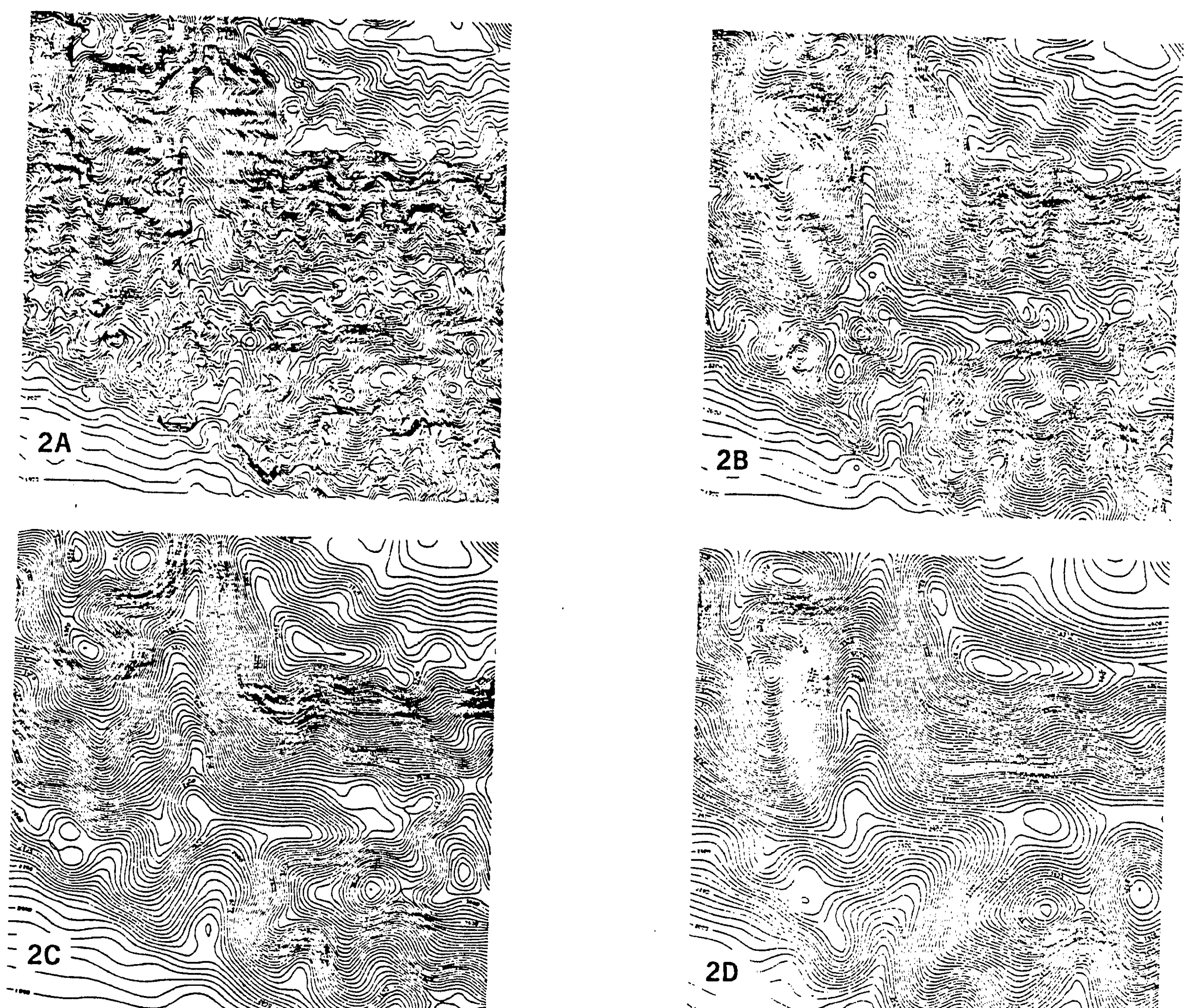


Figure 2—Computer generated plots of digitized topography for the area shown in Figure 1. The plots are based on elevation values at gridpoints on a 400-foot (122-m) mesh. The gridpoint data represents actual values (2A), values interpolated from an 800-foot mesh (2B), values interpolated from a 1200-foot mesh (2C), and values interpolated from a 1600-foot mesh (2D).

APPLICATION

The cubic spline may be used to interpolate data or to smooth data, to estimate volume or to replace or supplement finite difference approximations in numerical calculations. Before applying the spline technique for interpolation, the user must understand the nature of the available data and remember that the spline is a technique for achieving a polynomial fit with minimum curvature. If the actual data contain significant variations, or frequencies (wavelengths) that are higher than sample data collection, then the root mean square error associated with spline interpolation is not likely to be significantly better than the root mean square error associated with a linear interpolation scheme. The spline interpolation scheme will provide a minimum curvature fit, to the data, that is continuous through the second derivative. Spline interpolation is advisable, therefore, when the interpolated data are to be used in calculations which may be adversely affected by discontinuous slopes.

As an example of data interpolation and smoothing, we applied the bi-cubic spline technique to digital terrain heights on a grid. The topography of the Devil Canyon area of the San Bernardino National Forest in southern California (*fig. 1*) had been hand-digitized at a horizontal spacing of 400 feet (122 m) and an elevation resolution of 20 feet (6 m). A meteorological research network is located there and is used in the development of numerical weather models which also require topographic data. We created new data sets by using every other point (800-foot spacing) of the original grid, every third point (1200-foot spacing), and every fourth point (1600-foot spacing). The new data sets were then interpolated by bi-cubic splines to the original grid points. The root mean square errors in elevation at the interpolated points that resulted were 67 feet at the 800-foot mesh length, 93 feet at 1200 and 119 feet at 1600. *Figures 2a, b, c, and d* show computer-generated plots of the original 400-foot digitization and the three interpolations. Note how the many small ravines of 500-foot to 1000-foot width that are seen in the quadrangle map (*fig. 1*) are represented and smoothed in the four plotted digital maps (*fig. 2*). If it is assumed that a grid of mesh length Δ can resolve features with a 2Δ characteristic scale, then one can expect features of 800, 1600, 2400, 3200-foot widths to be resolved in *figs. 2a, b, c, and d.* The resolutions we achieved were actually somewhat better. Major topographic features such as the east and west forks of Devil Canyon, the Pine Flat and Camp Paivika plateaus, and Monument Peak are discernible through the 1600-foot mesh (*fig. 2d*).

Users should determine the balance between resolution and smoothing required for their particular needs and choose an appropriate interpolation interval. If variable resolution is required, the spline can telescope to higher resolution in areas of interest while maintaining lower resolution elsewhere. This technique can be applied in digitizing the data by choosing a tight digitizing interval where gradients are large and allowing looser intervals elsewhere, and then having the variable mesh length spline generate an equal mesh length grid.

A FORTRAN IV computer program, written by Julio L. Guardado, University of California, Riverside, will yield a bi-cubic spline interpolation for either a variable mesh length or an equal interval grid. Copy of the program is available on request to: Director, Pacific Southwest Forest and Range Experiment Station, P.O. Box 245, Berkeley, California 94701, Attention: Computer Services Librarian.

NOTES

[1] Fritsch, J. M. 1971. *Objective analysis of a two-dimensional data field by the cubic spline technique.* Month. Weather Rev. 99:379-386.

[2] Price, G. V., and A. K. MacPherson 1973. *A numerical weather forecasting method using cubic splines on a variable mesh.* J. Appl. Meteorol. 12:1102-1113.

[3] Simmonds, I. 1975. *On interpolation and evaluation of derivatives from a finite number of equally-spaced data points.* J. Appl. Meteorol. 14:1004-1010.

[4] Ahlberg, J. H., E. N. Nilson, and J. L. Walsh. 1967. *The theory of splines and their applications.* Mathematics in Science and Engineering, Vol. 38. 284 p. Academic Press, New York and London.

The Author

WILLIAM T. SOMMERS, a research meteorologist, is assigned to the Station's forest fire meteorology research unit, at the Forest Fire Laboratory, Riverside, California. He earned degrees in meteorology at City College of New York (B.S., 1965), Massachusetts Institute of Technology (M.S., 1967), and State University of New York at Albany (Ph.D., 1972). He joined the Forest Service and Station staff in 1973.

PACIFIC SOUTHWEST
Forest and Range
Experiment Station

FOREST SERVICE
U.S.DEPARTMENT OF AGRICULTURE
P.O. BOX 245, BERKELEY, CALIFORNIA 94701

Interpolation of unevenly spaced data using a parabolic leapfrog correction method and cubic splines

Julio L. Guardado William T. Sommers

USDA Forest Service
Research Note PSW-324
1977

Data collection sites used during field experiments or in operational observing networks are usually not located at points of an orthogonal grid. But numerical analysis techniques for processing the collected observations usually require data on an orthogonal grid. Even when data from unevenly spaced points are interpolated to other unevenly spaced points, many techniques require an intermediate orthogonal grid.

In this note, we describe a technique by which data recorded at unevenly spaced sites are used with cubic splines[1] to interpolate to unevenly spaced sites or to generate an orthogonal grid. The technique may be applied to any type of data. Judicious choice of grid spacing will, however, improve performance when the general variation of the specific data being interpolated is known.

PROCEDURE FOR INTERPOLATING

This procedure for interpolating from unevenly spaced data has three steps. First, we use the known data to generate an initial "guess field" grid by means of a simple square-of-distance weighting scheme using straight-line distance between points. Then, we adjust the initial guess field rows and columns using the known data points and a parabolic leapfrogging technique. Finally, we interpolate the corrected field to a desired orthogonal grid or to another set of unevenly spaced points.

Initial guess field rows and columns are chosen such that each row and column passes through one of the known data points. Normally, only one column is generated per unique data point x value and only one row is generated per unique data point y value. The grid produced is orthogonal but has unevenly spaced rows and columns. After the initial rows and columns are adjusted, cubic spline interpolation[1] is performed

Guardado, Julio L., and William T. Sommers
1977. Interpolation of unevenly spaced data using a parabolic leapfrog correction method and cubic splines. USDA Forest Serv. Res. Note PSW-324, 5 p., Pacific Southwest Forest and Range Exp. Stn., Berkeley, Calif.

The technique proposed allows interpolation of data recorded at unevenly spaced sites to a regular grid or to other sites. Known data are interpolated to an initial guess field grid of unevenly spaced rows and columns by a simple distance weighting procedure. The initial guess field is then adjusted by using a parabolic leapfrog correction and the known data. The final output is then generated by use of cubic spline interpolation on the adjusted grid. Application to test and actual data showed acceptably low root-mean-square error.

Oxford: 111.0–015.5.
Retrieval Terms: meteorological analysis; data processing; cubic splines.

independently on each row and column to generate two data values at the desired output points. These two data values are then averaged to produce the final output.

Inverse Square-of-Distance Weighting

The problem of determining a function value at a random point on a grid was solved using an inverse square-of-distance weighting applied to the known point function values. The basic procedure for finding a value at an unknown point from weighted values at n known points is

$$f' = \frac{\sum_{i=1}^{n} w_i f_i}{\sum_{i=1}^{n} w_i}$$

where f' is the function value at the unknown point, f_i are the function values at the n known points, and w_i are the weights given to the f_i. Because we desire that the contributions of the f_i to f' be weighted as the inverse square of the distance, then

$$w_i = \frac{1}{d_i^2}$$

where d_i are the distances between known points and unknown point. This procedure is repeated for each unknown point on the row or column of the initial guess grid which we are filling.

For example, say that the point x_1 was a distance d_1 from a point x' at which we would like to approximate the function value. Furthermore, the point x_2 is twice as far from x' as x_1. Point x_1 would be given a weight of 4/5 and point x_2 would receive a weight of 1/5 because it is twice as far away as x_1 and should be given 1/4 the weight of x_1. The sum of the weights multiplied by the function values at the known points yields the function value at the unknown point.

The Parabolic Leapfrog

Since all the initial guess rows and columns pass through at least one known data point, the question "How can we correct the initial guess given a known point?" arises. It was decided to correct the initial rows and columns by a "leapfrog" technique beginning at the known data point and performing piecewise corrections using a parabola passed through three points to obtain a fourth point within that interval. The leapfrogging procedure is diagrammed for a case where (a) leftmost endpoint is known, (b) rightmost endpoint is known, and (c) a "central" point is known. The points marked with an asterisk are used to determine those marked with a question mark. The numbers signify the leapfrogging sequence.

(a) (+) + + + +
1 * * ? *
2 * * ? *

(b) + + + + (+)
* ? * * 1
* ? * * 2

(c) + + (+) + +
* * ? * 1
* ? * * 2

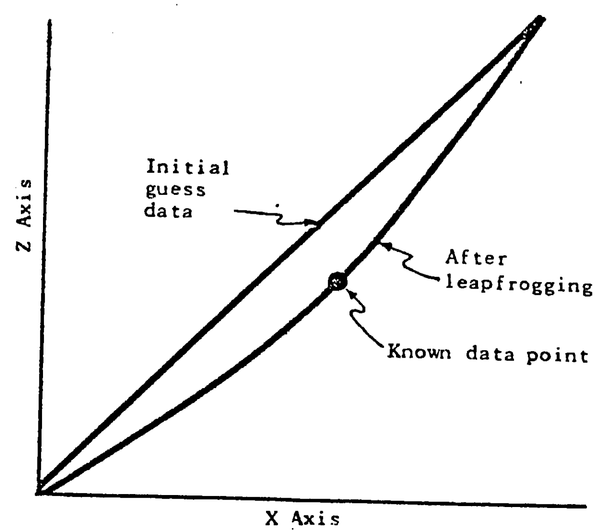


Figure 1—The curve demonstrates the effect of parabolic leapfrogging applied to a straight line and a known data point not on that line.

Figure 1 shows the effect of the leapfrogging procedure on a straight line, given a known point not on the line. We can see by the distance between the two lines that the effect of the correction procedure diminishes as the distance from the known point increases.

Given a set of points

$$(x_i, f_i) \text{ for } i = 0,1,2,3,\ldots n \quad n \geq 3$$

and given a known point (x_k, f_k) $0 \leq k \leq n$, then the remaining f_i $(i = 0, \ldots n \quad i \neq k)$ may be corrected by using the technique of parabolic leapfrogging.

Let j be the starting subscript for choosing the points to be used for interpolation and correction in a negative (decreasing x_i) direction, and let m be the starting subscript for interpolation and correction in a positive (increasing x_i) direction. We must consider three cases when choosing j and m.

Case 1: $x_k = x_n$, then $j = n$ and $m = \theta$ (not defined)

Case 2: $x_k = x_0$, then $j = \theta$ and $m = 0$

Case 3: x_k $k \neq 0,n$, then $j = \min(n,k+1)$ and $m = \max(0,k-1)$

We perform the parabolic leapfrog corrections in a positive direction by choosing the set of points

$$\{(x_i,f_i), (x_{i+1},f_{i+1}), (x_{i+3},f_{i+3})\}$$

where $m+3 \leqslant n$ and $m \leqslant i \leqslant n-3$, to find the point (x_{i+2},f_{i+2}).

For leapfrogging in the negative direction we choose the set of points

$$\{(x_{i-3},f_{i-3}), (x_{i-1},f_{i-1}), (x_i,f_i)\}$$

where $j \geqslant 3$ and $j \geqslant i \geqslant 3$, to find the point (x_{i-2},f_{i-2}).

We note from the above that at least four points are required for this correction procedure and that the leapfrogging cannot always be done in both a positive and negative direction. Leapfrogging could not be used to correct a boundary value but can be used on interior points nearest to the boundaries. Because interpolation, by definition, requires known data to bound unknown points, correcting boundary values should not be required.

The Lagrange Interpolating Polynomial

A parabola was fitted through the set of "known" points in the leapfrogging stage using a procedure known as Lagrange interpolation. This technique was chosen because of its ease of use and its flexibility. A polynomial of degree n-1 is fitted through n points, so that if desired, one could include more points in the leapfrog's individual interpolations with minor modifications to the procedure.

Let

$$P(x) = \sum_{i=0}^{n-1} f_i L_i(x)$$

where
- $P(x)$ = the interpolated value at x
- f_i = the value of the function to be approximated at known point x_i
- $L_i(x)$ = the i^{th} Lagrange coefficient at x
- x_i = the i^{th} known x value
- n = the number of interpolating points

We define the Lagrange coefficient as

$$L_k(x) = \prod_{\substack{m=0 \\ m\neq k}}^{n-1} \frac{x-x_m}{x_k-x_m}$$

for $k = 0,1,\ldots, n-1$.

Henrici[2] gives the motivation for this technique and also a proof.

Cubic Splines

When a grid has been generated and corrected, cubic spline interpolation is used to supply data values at the desired unknown points. The cubic spline is a cubic polynomial that can be used with data sets of either fixed or variable mesh length. Cubic spline interpolation provides a minimum curvature fit to the data that is continuous through the second derivative and returns exact values at known points. For additional details see Sommers (1976).[1]

APPLICATION

We first tested the grid generation scheme by reconstructing an existing grid of 51 rows and 51 columns with a data field consisting of a conical surface centered on gridpoint (26,26). The cone's surface formed a 30° angle with the plane of the grid and the data values ranged from 0.0 at (26,26) to slightly over 20.3 at the grid's corner points. The "known" points for the grid interpolation were chosen at random from every row, every other row, and every third row. The root-mean-square errors of the unknown points in each case are summarized in *table 1*. A FORTRAN IV program compiled by the FORTRAN H optimizing compiler and executed on an IBM 360 model 50[3] was used to perform the interpolation from the known points to the desired 51 X 51 grid. Execution times for each run are included in *table 1* to give a feel for percentage change when the number of points is increased or decreased. The ratio of execution times would be fairly constant for most machines.

Table 1—*Error and execution time for interpolation on a conical surface to a 51 x 51 grid*

Number of known points	Root-mean-square error[1]	Execution time (*min*)
51	1.5179	1.53
26	2.5315	0.53
17	3.1588	0.46

[1] Root-mean-square errors were determined by taking the values at known points, dropping them out, interpolating to them as if they were unknown, and comparing the results.

As a second test, we selected a set of 71 standard reporting weather stations in the Southeastern United States, held out a subset as the "unknown" points, and interpolated the "known" points to the unknown points by generating a grid, using the same procedure as in the first test case, and then interpolating the gridded data to the unknown points. *Figure 2* shows the location of the 50 known points and the 21 unknown points. We performed the interpolation procedure on three variables: temperature, sea level temperature, and dewpoint.

During the initial attempt at interpolating the real-world data we discovered that the "shooting" error which sometimes arises when one is fitting curves became quite large and unacceptable. Shooting errors arise when the gradient between two known points is so large that the interpolating function trajectory continues to increase or decrease beyond reasonable limits at an unknown point downstream. The shooting error was corrected two different ways for comparison. The first and most obvious way to correct it was to limit the interpolated value to either the minimum or the maximum of the points being used for interpolation according to whether we undershot or overshot. The second was to fit the point using linear interpolation if shooting occurred. We ran the program both ways for six hourly observations between 1800 and 2300 GMT on April 16, 1976, and computed root-mean-square errors at each of the "unknown" points. Linear interpolation resulted in minimum root-mean-square errors in all cases analyzed. The results are given in *table 2.*

Table 2—*Root-mean-square temperature, sea level temperature, and dewpoint errors at interpolated stations for hour observations between 1800 and 2300 GMT on April 16, 1976, using two methods for correcting shooting error*

Correction method	Root-mean-square error for interpolated . . .		
	Temperature	Sea level temperature	Dewpoint
	°C		
Minimum/maximum limiting	2.46	2.51	3.76
Linear interpolation	2.42	2.43	3.38

The points used as the "known" points were chosen at random and the grid used for the interpolation was arbitrarily given 25 rows and columns. This grid size was chosen to hold computer core and cpu requirements down in an effort to speed up the debugging process by increasing the total number of runs per day. A careful analysis of the known data available and a judicious choice of the grid's mesh length based on the user's particular data base should

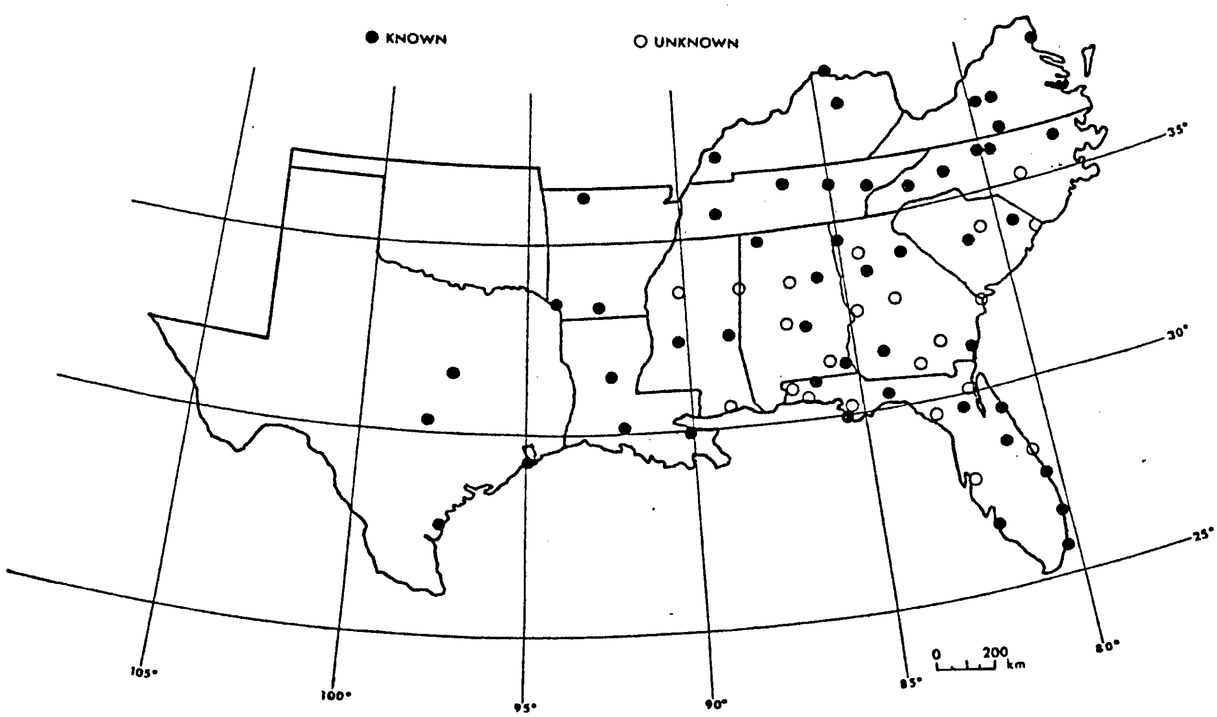


Figure 2—Meteorological data from "known" sites in the 13 southeastern States were interpolated to "unknown" sites in the six most southeasterly States.

yield even better results. These two considerations are the most important when using this particular scheme for interpolation of data.

FORTRAN IV routines which will perform the interpolation of randomly spaced data to grid, and which will interpolate gridded data to randomly spaced points, are available on request to: Director, Pacific Southwest Forest and Range Experiment Station, P.O. Box 245, Berkeley, California 94701, Attention: Computer Services Librarian.

NOTES

[1] Sommers, W. T. 1976. *Data interpolation by cubic splines.* USDA Forest Serv. Res. Note PSW-313, 5 p. Pacific Southwest Forest and Range Exp. Stn., Berkeley, Calif.

[2] Henrici, P. 1964. *Elements of numerical analysis.* p. 183-193. John Wiley and Sons, New York.

[3] Trade names and commercial enterprises or products are mentioned solely for information. No endorsement by the U.S. Department of Agriculture is implied.

The Authors

JULIO L. GUARDADO, an assistant programmer, performed the work reported here while assigned to the Station's forest fire meteorology research unit, at the Forest Fire Laboratory, Riverside, California. He earned a B.S. degree in applied science at the University of California, Riverside in 1976. He is presently a programmer-analyst with California Computer Products (CalComp), Anaheim, Calif.

WILLIAM T. SOMMERS, a research meteorologist, is assigned to the forest fire meteorology research unit, at Riverside, California. He earned degrees in meteorology at City College of New York (B.S., 1965), Massachusetts Institute of Technology (M.S., 1967), and State University of New York at Albany (Ph.D., 1972). He joined the Forest Service and Station staff in 1973.

☆U. S. GPO: 1978—793-121

Attachment Q

COMPARING METHODS OF FILLING IN MISSING DATA

Suppose one has a number of years of monthly average values of pollutant readings, and that in some one year, some of the monthly averages are missing. Number the months by $i = 1,2,\ldots,12$. We will address our remarks to three different methods of filling in the missing values.

The first assumes that the seasonal pattern is fairly constant. Thus, if X_{ij} is the i^{th} monthly average

$$X_{ij} = P_i\ T_j(1+E_{ij})$$

where T_j is the total for the j^{th} year, $\sum_i P_i = 1$, and as a consequence

$$T_j = \sum_i X_{ij} = T_j \sum_i P_i(1+E_{ij})$$

implies that the random errors E_{ij} satisfy

$$(1) \qquad\qquad \sum_i P_i\ E_{ij} = 0, \text{ all } j.$$

We will assume that the E_{ij} are normally distributed with equal variances σ^2 and mean zero conditioned on (1).

Suppose that we have J years of good data, and in year $j = o$, we have good data in the months $I = \{i_1,\ldots,i_m\}$ and data missing in the other months whose set of indices we denote by I^c. A natural method of filling in the missing data is to estimate the P_i from the J good years by taking

$$\hat{P}_i = \frac{1}{J} \sum_{j>0} X_{ij}/T_j$$

$$= \frac{1}{J} \sum_{j>0} P_i(1+E_{ij}) = P_i\left(1+\frac{1}{J}\sum_{j>0} E_{ij}\right) .$$

Assume also that the E_{ij} are independent for different j and independent for i before the conditioning (1).

Now estimate the total T_o for the year with missing data by

$$\hat{T}_o = \frac{1}{I} \sum_{i\varepsilon I} X_{ij}/\hat{P}_i$$

$$= \frac{1}{I} \sum_{i\varepsilon I} P_i \; T_o(1+E_{io})/\hat{P}_i$$

$$= T_o\left(\frac{1}{I} \sum_{i\varepsilon I} (1+E_{i,o})P_i/\hat{P}_i\right) .$$

Now fill in the missing data by the estimate

$$\hat{X}_{io} = \hat{P}_i \; \hat{T}_o .$$

We want to estimate the size of the overall bias over the missing months. This bias is defined as

$$B = \sum_{i\varepsilon I^c} (\hat{X}_{io} - P_i \; T_o).$$

As our measure of bias, we will compute EB^2. Now

$$B = \sum_{i\varepsilon I^c} \hat{x}_{io} - T_o \sum_{i\varepsilon I^c} P_i \quad .$$

Now

$$\hat{x}_{io} = P_i \ T_o\left(1+ \frac{1}{J} \sum_{j>o} E_{ij}\right)\left(\frac{1}{I} \sum_{i\varepsilon I} (1+E_{i,o})P_i/\hat{P}_i\right).$$

Write

$$\frac{1}{I} \sum_{i\varepsilon I} (1+E_{io})P_i/\hat{P}_i = 1 + \frac{1}{I} \sum_{i\varepsilon I} \frac{P_i-\hat{P}_i}{\hat{P}_i} + \frac{1}{I} \sum_{i\varepsilon I} E_{io} \ P_i/\hat{P}_i \quad , \tag{2}$$

and approximate

$$\frac{P_i-\hat{P}_i}{\hat{P}_i} \simeq \frac{P_i-\hat{P}_i}{P_i} = - \frac{1}{J} \sum_{j>o} E_{ij}$$

$$\sum_{i\varepsilon I} E_{io} \ P_i/\hat{P}_i \simeq \sum_{i\varepsilon I} E_{io}$$

to get the expression in (2) to be

$$1 - \frac{1}{J\,I} \sum_{\substack{j>o \\ i\varepsilon I}} E_{ij} + \frac{1}{I} \sum_{i\varepsilon I} E_{io}$$

Therefore, to first-order terms

$$\hat{x}_{i,o} = P_i \ T_o\left(1+\frac{1}{J} \sum_{j>o} E_{i,j} - \frac{1}{IJ} \sum_{\substack{j>o \\ i\varepsilon I}} E_{ij} + \frac{1}{I} \sum_{i\varepsilon I} E_{io}\right)$$

so

$$\hat{x}_{i,o} - P_i \ T_o = P_i \ T_o\left(\frac{1}{J} \sum_{j>o} E_{i,j} - \frac{1}{IJ} \sum_{\substack{j>o \\ i\varepsilon I}} E_{ij} + \frac{1}{I} \sum_{i\varepsilon I} E_{io}\right).$$

Summing over $i \varepsilon I^c$ gives

$$B/T_o \simeq \frac{1}{J} \sum_{\substack{j>o \\ i\varepsilon I^c}} P_i\ E_{i,j} - \frac{P}{I} \sum_{\substack{j>o \\ i\varepsilon I}} E_{ij} + \frac{P}{I} \sum_{i\varepsilon I} E_{io}$$

where $P = \sum_{I^c} P_i$. Since $\sum_i P_i\ E_{ij} = 0$, we can write

$$\sum_{\substack{j>o \\ i\varepsilon I^c}} P_i\ E_{ij} = - \sum_{\substack{j>o \\ i\varepsilon I}} P_i\ E_{ij}$$

and rewrite

$$B/PT_o \simeq - \frac{1}{J} \sum_{\substack{j>o \\ i\varepsilon I}} E_{ij} \left(\frac{Pi}{P} + \frac{1}{I}\right) + \frac{1}{I} \sum_{i\varepsilon I} E_{io} .$$

Note that the two terms are independent, and have mean zero.

We can go a little further using a standard probability result, i.e., let $Z_1, \ldots, Z_n$ be independent $N(o,\sigma^2)$ variables. Then

$$\text{(3)} \qquad E\left[\left(\sum_j \alpha_j\ Z_j\right)^2 \,\middle|\, \sum_j \beta_j\ Z_j = 0\right] = \sigma^2 \left[\sum_j \alpha_j^2 - \frac{\left(\sum \alpha_j \beta_j\right)^2}{\sum \beta_j^2}\right] .$$

From the above, we have that

$$\text{(4)} \qquad \frac{1}{P^2 T_o^2}\ EB^2 = \frac{1}{J^2} E \left[\sum_{\substack{j>o \\ i\varepsilon I}} E_{ij} \left(\frac{P_i}{P} + \frac{1}{I}\right)^2 \,\middle|\, \sum_i P_i\ E_{ij} = 0\right]$$

$$+ \frac{1}{I^2} E \left[\left(\sum_{i\varepsilon I} E_{io}\right)^2 \,\middle|\, \sum_i P_i\ E_{io} = 0\right] .$$

Since the E_{ij} are independent for different j, the first term is simply

$$\frac{1}{J} E\left\{\left[\sum_{i\varepsilon I} E_{i1}\left(\frac{P_i}{P} + \frac{1}{I}\right)\right]^2 \middle| \sum_i P_i E_{i1} = 0\right\} .$$

By (3) above this equals

$$(5) \qquad \frac{\sigma^2}{J}\left\{\sum_{i\varepsilon I}\left(\frac{P_i}{P} + \frac{1}{I}\right)^2 - \frac{\left[\sum_{i\varepsilon I} P_i\left(\frac{P_i}{P} + \frac{1}{I}\right)\right]^2}{\sum_i P_i^2}\right\} . .$$

Using (3) again, the 2^{nd} term in (4) is

$$(6) \qquad \frac{1}{I^2}\sigma^2\left[I - \frac{\left(\sum_I P_i\right)^2}{\sum P_i^2}\right] .$$

As a first approximation, take all $P_i = 1/12$. Then (5) becomes

$$\frac{\sigma^2}{J}\left(\frac{1}{I} + \frac{1}{I^c}\right)^2\left[I - \frac{I^2}{12}\right]$$

$$= \frac{\sigma^2}{J} \cdot \frac{12}{II^c} .$$

Also (6) becomes

$$\frac{\sigma^2}{I^2} \cdot \frac{II^c}{12} = \sigma^2 \frac{I^c}{12I}$$

Noting that $P = I^c/12$ we get

$$EB^2 = T_o^2 \sigma^2\left[\frac{12}{JII^c} \cdot \frac{(I^c)^2}{12^2} + \frac{I^c}{12I} \cdot \frac{I_c^2}{12^2}\right]$$

$$= \frac{\sigma^2 T_o^2}{12}\left[\frac{I^c}{JI} + \frac{1}{12^2}\frac{I_c^3}{I}\right]$$

(7) $$= \frac{\sigma^2 T_o^2}{12} \left(\frac{I_c}{I}\right) \left[\frac{1}{J} + \left(\frac{I^c}{12}\right)^2\right] .$$

On the other hand, suppose we fill in the missing values by averaging two adjacent years, i.e., take

$$\hat{x}_{i,o} = (x_{i,1} + x_{i,2})/2$$

$$= P_i \; [(T_1 + T_2)/2] + \frac{1}{2} P_i \; T_1 \; E_{i,1} + \frac{1}{2} P_i \; T_2 \; E_{i,2} \; .$$

Then the bias B is given by

$$B = \sum_{I^c} (\hat{X}_{i,o} - P_i \; T_o)$$

$$= \sum_{I^c} P_i \left(\frac{T_1+T_2}{2} - T_o\right) + \frac{1}{2} \sum_{i \varepsilon I^c} (T_1 P_i E_{i,1} + P_i T_2 E_{i,2}) \; .$$

Assuming again that $P_i = 1/12$, we get

$$EB^2 = \left(\frac{I^c}{12}\right)^2 E\left(\frac{T_1+T_2}{2} - T_o\right)^2 + \left(T_1^2 + T_2^2\right) \cdot \frac{\sigma^2}{12^2} \; \frac{II^c}{12} \cdot \frac{1}{4} \; .$$

If we take $T_1 = T_2 = T_o$ to give a lower bound, then we have

(8) $$EB^2 \geq \frac{1}{2} T_o^2 \sigma^2 \frac{II^c}{12^3} \; .$$

A simple calculation shows that the expression in (7) is lower than the bound in (8) for $I^c \leq 3$, for all $J \leq 8$. In other words, for 3 or less months of missing data, if $J \leq 8$, use the first method. If one assumes that $\left(\frac{T_1+T_2}{2} - T_0\right)^2$ is small, then for 4 or more months of missing data, use the 2_{nd} method.

If we only use one adjacent year to fill in, the expression for EB^2 is

$$EB^2 = \left(\frac{I^c}{12}\right)^2 E(T_1-T_0)^2 + T_1^{\,2} \frac{\sigma^2}{12^2} \cdot \frac{II^c}{12} \quad .$$

The lower bound with $T_1 = T_0$ is

(9) $$EB^2 \geq T_0^{\,2} \sigma^2 \frac{II^c}{12^3} \quad .$$

In this case, for J = 5, use the first method if 4 or less months are missing. For J = 7, it becomes 5 or less.